



Etudes biophysiques du facteur de maturation 3' des ARN pré-messagers CF IA

Adrien Dupin

► To cite this version:

Adrien Dupin. Etudes biophysiques du facteur de maturation 3' des ARN pré-messagers CF IA. Biochimie [q-bio.BM]. Université de Bordeaux, 2014. Français. NNT: 2014BORD0189 . tel-01272732

HAL Id: tel-01272732

<https://theses.hal.science/tel-01272732>

Submitted on 11 Feb 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Etudes biophysiques du facteur de maturation 3' des ARN pré-messagers CF IA

Thèse présentée et soutenue publiquement :

Le 6 novembre 2014

Par :

Adrien Dupin

Pour l'obtention du diplôme de :

Docteur de l'Université de Bordeaux

Ecole Doctorale : Sciences de la Vie et de la Santé

Spécialité : Interface Chimie – Biologie

Composition du jury :

M. Romier Christophe	Directeur de recherche	Strasbourg	Président Rapporteur
M. Petosa Carlo	Directeur de recherche	Grenoble	Rapporteur
Mme. Millevoi Stefania	Chargé de recherche	Toulouse	Examineur
M. Graille Marc	Directeur de recherche	Paris	Examineur
M. Fribourg Sébastien	Directeur de recherche	Bordeaux	Directeur de thèse

Remerciements

Sébastien, merci pour la relation que nous avons pu avoir durant ces trois ans, au-delà de la seule dimension professionnelle. Tu es un homme aux centres d'intérêts variés, ce qui nous a permis de partager beaucoup de choses. Le sport, avec son lot de coups de raquette, balles dans le dos ou même orteil cassé ! Ou encore les discussions de 8h du matin, de 8h du soir ou durant les missions au synchrotron ou les congrès, à propos de la vie de labo, de la vie perso, de l'actualité ou le plus souvent d'un documentaire entendu à la radio ou la télé. Mais au-delà de ça, tu es aussi une personne dont j'admire beaucoup l'éthique de travail, l'esprit de rigueur et l'importance que tu attaches au travail de paillasse. Ce sont des choses qui t'ont définitivement permis de gagner mon estime, et ce sont des comportements que j'espère arriver à me fixer comme objectifs tout au long de ma carrière. Humainement, j'ai toujours apprécié ta sollicitude et ta disponibilité, même s'il faut savoir apprivoiser tes remarques parfois cinglantes ! ☺ J'ai beaucoup appris à ton contact et même si j'avance maintenant vers beaucoup d'inconnu, je sais que ton apport me sera toujours des plus utiles. Merci.

Je ne peux pas non plus imaginer ce qu'aurai été une thèse sans les nombreux amis du labo. Pour moi, cela a toujours fait partie de ma représentation de la vie dans la recherche. Fanny & Lionel, Heddy & Jon, Britta & Carolin, Kishore & Axel, Denis & Cameron, Julia & Camilla, Guillaume & Gavin, Patricia & Patricia. Impossible de citer tous les gens que j'apprécie au sein de l'institut, ceux qui sont encore là, ceux qui sont partis et ceux qui n'étaient que de passage. Merci à tous pour les soirées plages, les soirées tout court, les discussions de couloir ou les conseils.

De la même manière, merci à tous mes amis de longue date pour leur support indéfectible durant ces trois ans ! Ça n'a pas toujours été évident, surtout de devoir décliner soirée après soirée pendant ma dernière année ! Ou encore mettre en suspend mon engagement au Chantier, mais tout le monde a été plus que compréhensif ! Il faut aussi parler des rendez-vous de soutien psychologique du lundi soir au Dick, thérapie de groupe entre thésards qui ont été un accompagnement de qualité !

Enfin, merci à ma famille. Il n'y aura jamais assez de mots pour ce que vous m'avez apporté, m'apportez et m'apporterez, ni pour combien je tiens à vous. Merci pour tout.

Sommaire

1. Données bibliographiques	1
1.1. Historique de la maturation 3'	2
1.2. La polyadénylation et ses mécanismes biochimiques.....	5
1.3. La pA chez la levure	6
1.3.1. Élément <i>cis</i>	6
1.3.2. Éléments <i>trans</i>	8
1.3.3. Focus sur CF IA.....	15
1.4. Comparaison machinerie Levure / Mammifères.....	21
1.5. Les autres maturations de l'extrémité 3'	22
1.6. Polyadénylation alternative et sélection du site poly(A)	24
1.7. Couplage avec les autres mécanismes	25
1.7.1. Transcription & Terminaison.....	25
1.7.2. La coiffe 5'	29
1.7.3. L'épissage	29
1.8. Maladies	30
2. Synthèse et objectifs.....	32
3. Résultats.....	33
3.1. Etude des interactions du CF I	33
3.1.1. La reconstitution du CF IA en modèle recombinant et son application pour la cartographie des interactions.....	33
3.1.2. Caractérisation des interactions <i>in vitro</i>	40
3.2. Etude biochimique de CF IA.....	46
3.2.1. Stratégies & constructions	46
3.2.2. Expression et purification.....	48

3.2.3.	Cristallisation	56
3.2.4.	Approche dynamique par la diffusion aux petits angles.....	68
3.2.5.	Microscopie électronique	77
3.2.1.	Spectrométrie de masse en conditions natives	94
3.3.	Cristallisation de CF I	117
3.3.1.	Quels partenaires pour stabiliser le CF IA	117
3.3.2.	Essais de cristallogenèse	120
3.4.	Etude des motifs en « doigts de zinc » de Pcf11p	121
3.4.1.	Analyse in silico	121
3.4.2.	Absorption atomique et liaison de zinc	124
3.4.1.	Analyse en MS	126
3.4.2.	Mutants Test in vivo & in vitro	129
3.4.3.	Essais de cristallisation	133
3.5.	Mise en évidence d'une région d'interaction étendue et conservée entre Cpl1p et Pcf11p	133
3.5.1.	La structure de Noble et les questions en suspend	133
3.5.2.	Les essais de cristallographie	134
3.5.3.	Article (PMID : 24508575)	134
4.	Conclusions et discussion	144
5.	Matériel et méthodes	149
5.1.	Biologie moléculaire	149
5.1.1.	Bactéries utilisées.....	149
5.1.2.	Plasmides utilisés	149
5.1.3.	Extraction d'ADN génomique de <i>Kluyveromyces lactis</i>	150
5.1.4.	Clonage	151
5.1.5.	Expression des protéines recombinantes	152

5.1.6.	Gel retard	153
5.1.7.	Purification de l'ARN polymérase du phage T7 P266L.....	153
5.1.8.	Transcription in vitro	154
5.1.9.	Western Blot.....	155
5.1.10.	Protéolyse ménagée	155
5.2.	Biologie	156
5.2.1.	Test en goutte et thermosensibilité.....	156
5.2.1.	Test de croissance	156
5.2.2.	Tests d'activité.....	156
5.3.	Biochimie	160
5.3.1.	Purification des protéines	160
5.3.2.	Purification des ARN.....	161
5.3.3.	Cristallogénèse	162
5.3.4.	GraFix.....	163
5.3.5.	Crosslink des protéines pour cristallogénèse.....	164
5.4.	Biophysique	164
5.4.1.	SAXS.....	164
5.4.2.	Cristallographie aux rayons X.....	165
5.4.3.	Microscopie électronique	165
5.4.4.	Spectrométrie de masse	165
5.4.5.	Absorption atomique	166
5.4.6.	SPR.....	166
6.	Bibliographie	168

1. Données bibliographiques

L'organisation biologique cellulaire, telle que nous la connaissons du moins, a longtemps été résumée dans le dogme ADN (information) -> ARN (messenger) -> Protéine (fonction). Ce schéma est plus nuancé de nos jours, mais il a l'avantage de mettre en avant un des nombreux rôles que peut tenir une molécule ribonucléotidique ; celui de messenger de l'information génétique.

L'acide ribonucléique dit « messenger », élément essentiel de la vie cellulaire et porteur transitoire de l'information génétique, s'il est relativement identique chimiquement, est considérablement différent chez les eucaryotes et chez les procaryotes d'un point de vue fonctionnel. En effet, la présence d'introns dans le gène eucaryote entraîne une première étape de maturation co-transcriptionnelle, l'épissage. Un résidu 7-méthyl-guanosine est également rajouté en 5' de la molécule d'ARN, modification connue sous le nom de « coiffe 5' ». Le produit de l'ARN polymérase II doit enfin être modifié sur son extrémité 3', modifications regroupées sous le nom de maturation 3' du pré-ARNm.

La caractéristique de l'expression des gènes eucaryotes est que les transcrits précurseurs de l'ARNm (pré-ARNm) subissent une série de réactions complexes de maturation cotranscriptionnelles avant leur export vers le cytoplasme (Figure 1).

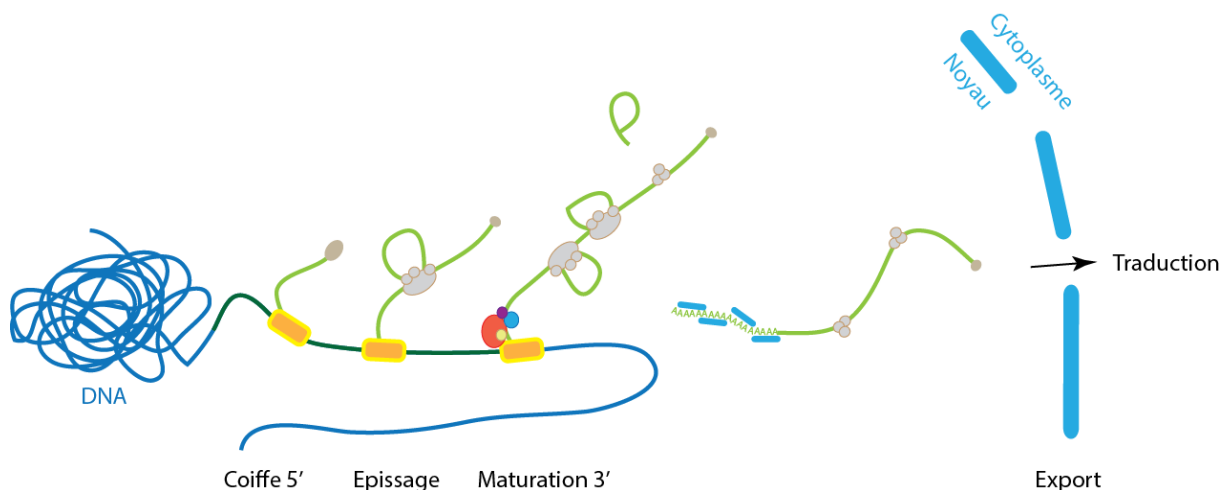


Figure 1 : Schéma du cycle de la vie nucléaire du transcrit eucaryote. L'ARN Pol II est représentée par le rectangle jaune.

Le pré-ARNm est synthétisé dans le noyau par l'ARN polymérase II (Pol II). Le premier événement de maturation à affecter le pré-ARNm fraîchement transcrit est le capping de son

extrémité 5' (Shuman 2001). Une fois que le transcrit naissant émerge de la Pol II, une ARN 5' triphosphatase (Cet1p chez la levure) convertit le triphosphate du premier nucléotide en diphosphate. Une guanylyl transférase fusionne alors un GMP aux phosphates terminaux via une liaison triphosphate 5'-5' et formant ainsi la coiffe (Shatkin and Manley 2000, Shuman 2001). La fonction primaire de la coiffe est de protéger l'ARNm d'une dégradation par des exonucléases 5'-3' (Beelman and Parker 1995) mais elle est également impliquée dans l'export et la traduction en permettant l'interaction spécifique avec certains facteurs (Gonatopoulos-Pournatzis and Cowling 2014).

Le pré-ARNm des eucaryotes contient des séquences, les introns, qui ne sont pas présentes dans l'ARNm mature. Les introns sont retirés au cours de la réaction d'épissage du pré-ARNm. L'épissage repose sur la reconnaissance d'un site d'épissage 5' AG|GURAGU et un site d'épissage 3' YAG|RNNN aux frontières exon-intron et intron-exon respectivement. La réaction d'épissage correspond à la connexion bout à bout des exons et l'excision des introns catalysée par le spliceosome, un important complexe riboprotéique (Jurica and Moore 2003), dont certains composants interagissent avec des facteurs de la maturation 3' (Proudfoot, Furger et al. 2002).

A l'exception des gènes des histones métazoaires dont l'expression est réplication-dépendante (Dominski and Marzluff 2007), tous les ARN messagers subissent une maturation 3' qui génère une queue poly(A) à l'extrémité 3' de chaque transcrit. La queue poly(A) peut être de différentes tailles (60 à 250 nt) selon l'organisme et l'ARNm (Sachs and Wahle 1993). La maturation 3' promeut la terminaison de la transcription, confère sa stabilité à l'ARNm, facilite l'export et la traduction (Colgan and Manley 1997, Richard and Manley 2009). La queue poly(A) est ajoutée par une réaction de clivage et de polyadénylation catalysée par la machinerie de maturation 3', un complexe multi-protéique qui reconnaît le signal poly(A) sur le transcrit naissant (Shi, Di Giammartino et al. 2009). Bien que ces réactions puissent être séparées biochimiquement *in vitro*, les événements de maturation du pré-ARNm sont interconnectés les uns avec les autres et avec la transcription dans la cellule, mais ils sont aussi liés à des événements postérieurs comme la traduction ou la dégradation de l'ARNm dans le cytoplasme (Moore and Proudfoot 2009, Perales and Bentley 2009).

1.1. Historique de la maturation 3'

L'histoire de l'étude de la polyadénylation a démarré il y a 55 ans, avec la découverte d'une enzyme qui polymérisait l'ATP en poly(A) (Edmonds and Abrams 1960), visiblement sans nécessiter de matrice ADN. De nombreuses autres Poly(A) polymérases furent mises en évidence dans d'autres organismes, mais leur existence ne prit un sens que grâce à la découverte de queues poly(A) sur les ARN messagers et leurs précurseurs nucléaires (Darnell, Wall et al. 1971, Edmonds, Vaughan et al. 1971, Lee, Mendecki et al. 1971). Rapidement, la queue poly(A) permit de grandes avancées techniques en autorisant la purification spécifique des ARNm ou en fournissant un site d'hybridation pour une amorce de rétro-transcription. Il fallut cependant attendre quelques années supplémentaires pour voir des avancées sur le rôle biologique de la queue poly(A) et de la région 3' non traduite, et beaucoup de ces résultats découlent de l'étude de l' α - et de la β -thalassémie et des gènes de globine.

Dans une série d'études de la seconde moitié des années 70, plusieurs preuves viennent mettre en évidence un rôle protecteur de la queue poly(A) qui protège l'ARNm de la dégradation *in vivo* (Huez, Marbaix et al. 1975, Huez, Marbaix et al. 1975, Huez, Marbaix et al. 1978). Dans un article publié en 1975 (Huez, Marbaix et al. 1975), Huez et al. micro-injectent des ARNm de globine radiomarqués dans des oocytes de grenouille et les extraient pour observer leur taille par centrifugation sur gradient de sucrose après environ 2h, 22h ou 46h. Alors que la présence des ARNm déadénylés n'est plus détectable à 22h, ils conservent un signal très fort pour les ARNm sur lesquels une queue poly(A) a précédemment été repolymérisée *in vitro*. En faisant varier le temps de polymérisation, ils parviennent également à mettre en évidence qu'un minimum de 30 résidus est nécessaire pour obtenir une stabilité identique à celle des ARNm natifs. De ces résultats, ils concluent que le segment poly(A) est responsable de la stabilité de l'ARNm.

En parallèle, Proudfoot et Brownlee ont montrés en 1976 (Proudfoot and Brownlee 1976) la forte conservation de séquence présente dans le 3'UTR de six ARNs messagers. A l'époque, alors que la région 3' non traduite est plus longue que celle en 5', aucun rôle biologique ne lui est encore attribué. Grâce à une nouvelle stratégie expérimentale, ils parviennent à séquencer les régions en amont de la queue poly(A) sur les ARNm des α - et β -globines de l'homme et du lapin ainsi que celle de l'immunoglobuline légère de la souris et l'ovalbumine de poulet. Ils mettent alors en évidence sur tous ces ARNm la présence d'une séquence conservée, AAUAAA, située 14 à 20 nucléotides en amont de la queue poly(A). Aucune fonction ne peut

encore y être associée, bien que sa proximité conservée avec le site de polyadénylation suggère fortement une implication dans les mécanismes de maturation 3'. Malgré des données de séquence parcellaires à l'époque, ils arrivent tout de même à aller plus loin dans leur réflexion en faisant une comparaison intéressante. En partant du postulat que l'importance fonctionnelle d'une séquence nucléotidique particulière est corrélée avec son degré de conservation entre les espèces, ils s'appliquent à comparer les conservations des séquences codante et non codante des gènes α - et β -globine entre l'humain et le lapin. Leurs observations sont les suivantes. Entre les deux espèces, le taux de mutations causant une substitution d'acide aminé dans la séquence codante est de 5,9%. Par contre, 55% de la séquence codante est sujette à des mutations neutres. Et comparativement, ils n'observent que 15% de divergence dans le 3' UTR. Bien que moins forte que celle s'appliquant sur la conservation de la séquence en acide aminés, il y a donc une pression de sélection significative qui impacte le 3'UTR, dont ils concluent qu'il doit porter une fonction séquence spécifique.

Deux ans plus tard, un outil de génétique innovant fut mis en place en clonant des fragments de restriction du génome de l'adénovirus, permettant notamment de suivre le devenir du transcrit tardif 25knt de l'adénovirus de type 2. Ce transcrit est issu d'un promoteur unique, mais génère cinq familles d'ARN polyadénylés, familles portant chacune un 3'UTR spécifique. La question autour de ce cas de polyadénylation alternative fut donc : comment est-ce que l'extrémité 3' est sélectionnée ? En 1978, Nevins et al. (Nevins and Darnell 1978) ont été capable de mesurer que la fréquence d'utilisation de chaque site de polyadénylation était de 10, 23, 28, 21 et 22%, par ordre d'apparition dans l'unité de transcription. Or, un choix aléatoire durant la transcription conduirait à une polarisation par laquelle le site proximal serait favorisé, puis la fréquence d'utilisation irait décroissante. Ce n'est pas le cas. Pour compenser cela, il doit donc exister des mécanismes de régulation dans les affinités de liaison des complexes de maturation et la sélection du site polyA

Les premiers essais de création d'un système *in vitro* utilisèrent des extraits nucléaires, ou des préparations à partir de cellules infectées par des adénovirus (Cooper and Marzluff 1978, Chen-Kiang, Wolgemuth et al. 1982), mais permettaient uniquement de tester le substrat pré-ARNm et non les facteurs protéiques, et ne permettaient pas non plus de découpler transcription, clivage et polyadénylation. Dans le début des années 1980, le test fut amélioré de manière significative, permettant d'exclure l'activité de transcription d'observer un clivage

très précis au site poly(A) (Manley 1983, Moore and Sharp 1985). Une nouvelle génération de tests *in vitro* fut développée chez la levure quelques années plus tard, permettant cette fois-ci de découpler les activités de clivage et de polyadénylation et de tester une grande variété de pré-ARNm (Butler and Platt 1988, Butler, Sadhale et al. 1990). L'amélioration des connaissances sur les facteurs protéiques amenée par ce test, couplée à des purifications de plus en plus sélectives (Chen and Moore 1992), a conduit à l'identification et la caractérisation progressive des facteurs protéiques impliqués dans la maturation (Takagaki, Ryner et al. 1989, Minvielle-Sebastia, Preker et al. 1994, Amrani, Minet et al. 1997, Minvielle-Sebastia, Preker et al. 1997, Dichtl, Blank et al. 2002).

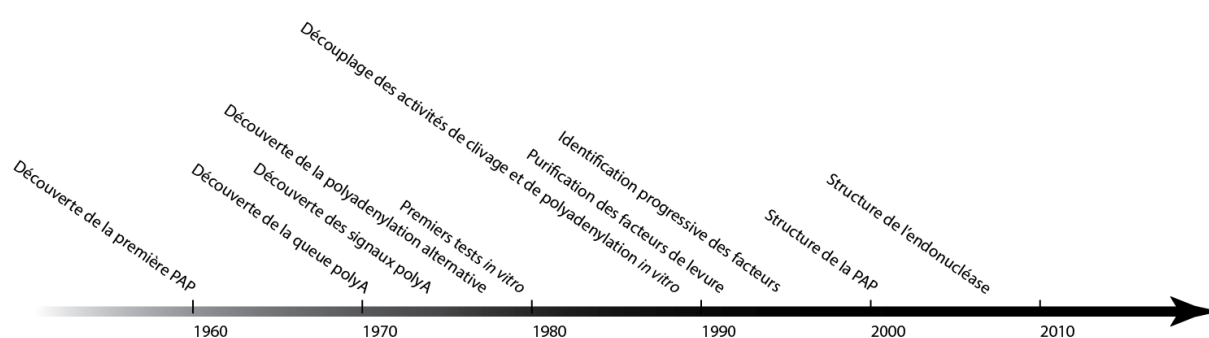


Figure 2 : Frise chronologique reprenant les principales avancées dans l'étude de la maturation 3'.

C'est dans cette vague que se situe le travail présenté dans ce document. Les avancées effectuées dans l'identification et les connaissances des facteurs protéiques posèrent rapidement la question de l'organisation de ces complexes. Des avancées ont été faites sur les protéines isolées, entières ou non, mais relativement peu d'interactions sont caractérisées structurellement, et peu de choses sont connues sur l'agencement des complexes dans l'espace.

1.2. La polyadénylation et ses mécanismes biochimiques

La maturation 3' des ARNm eucaryotes est une réaction en deux étapes : une endonucléase clive le pré-ARNm et une poly(A) polymérase synthétise une queue polyadénosine sur le fragment en 3' du fragment aval du site de clivage. Cette maturation co-transcriptionnelle est catalysée par un ensemble de facteurs protéiques organisés en complexes. Leurs rôles sont multiples : l'intégration des régulations, la reconnaissance spécifique du site de polyadénylation à traiter, le clivage endonucléotidique de la séquence d'ARN et la synthèse d'une séquence poly-adénine sur l'extrémité 3' créée.

L'endonucléase est une hydrolase zinc-dépendante de la superfamille des métallo- β -lactamases. La signature de cette famille est constituée de 5 motifs extrêmement conservés, responsables des interactions avec les ions Zn^{2+} du site actif. L'homologue humain est composé de 2 domaines. Le domaine N-terminal est le domaine canonique des métallo- β -lactamases alors que le centre de la protéine est plus spécifique, formant notamment une petite β -hairpin à la surface, spécifique de la sous-famille des β -CASP. La réaction nucléase est créée par l'attaque nucléophile d'un hydroxyde conservé contre un groupement phosphate du substrat d'ARN en interaction avec des atomes de zinc.

La polyA polymérase est une ARN polymérase matrice indépendante qui appartient à une famille de polymérases caractérisée par un motif « helix-turn ». Elle possède trois domaines globulaires. Un domaine central flanqué du domaine N-terminal catalytique, porteur du motif typique et un domaine C-terminal proche des domaines de reconnaissance de l'ARN. Les trois domaines participent à la liaison du substrat d'ARN en formant une crevasse capable d'interagir avec les trois derniers nucléotides de la séquence et un nucléotide triphosphate. Le site actif ainsi formé est extrêmement conservé entre la multitude de PAP existante, contrairement à la surface.

Rien n'est connu sur la coordination de ces deux activités enzymatiques lors de la maturation 3'. L'enzyme responsable du clivage, Ysh1p (CPSF-73 chez l'homme), n'a été formellement identifiée comme telle que récemment (Mandel, Kaneko et al. 2006) et nous sommes toujours loin d'avoir une image claire du déroulement de la maturation.

1.3. La pA chez la levure

1.3.1. Élément *cis*

Afin que la maturation 3' se déroule correctement, le pré-ARNm a besoin de séquences *cis* spécifiques pour guider les facteurs protéiques. Ces séquences sur le transcrit qui conditionnent le site de clivage et de polyadénylation sont englobées avec le terme de signal (ou signaux) poly(A). Le signal poly(A) est présent à l'extrémité 3' transcrite mais non traduite du gène.

La découverte qu'un mutant de délétion de 38nt dans le 3'UTR du gène *CYC1* résultait en la formation d'une extrémité 3' anormale est la première observation concernant les signaux poly(A) chez la levure (Zaret and Sherman 1982). Des expériences successives avec des constructions génétiques révélèrent à la fois le positionnement et la séquence de plusieurs signaux putatifs (Russo, Li et al. 1991). L'étude d'autres gènes de levure et du transposon Ty amenèrent à l'identification d'éléments riches en UA comme favorisant la maturation 3'. On parla d'efficiency element (EE). Des études additionnelles avec le système *CYC1* (Sherman 2005) ont ensuite établies l'importance d'un élément de positionnement (PE), localisé entre l'EE et le site poly(A) (Russo, Li et al. 1993). Des études mutationnelles montrèrent que le PE tolérait une dégénérescence significative, tandis que AAUAAA ou AAAAAA sont optimaux. Le site de clivage en tant que tel est peu conservé, mais se retrouve habituellement après une pyrimidine et quelques adénosines, représenté par le motif consensus PyA. Cette structure tripartite PE EE CS fut confirmée par des études menées sur un site de polyadénylation synthétique (Guo and Sherman 1996).

Des études bio-informatiques ultérieures à plus grande échelle analysant des centaines d'EST (expressed sequence tag) confirmèrent largement le positionnement et la conservation des éléments identifiés génétiquement et biochimiquement, mais révélèrent également l'existence de deux domaines additionnels, riches en U et situés de part et d'autre du site de clivage (Graber, Cantor et al. 1999, van Helden, del Olmo et al. 2000). Le rôle biologique de ces séquences fut rapidement confirmé *in vitro* sur des séquences de levure (Dichtl and Keller 2001).

Le modèle actuel du site de maturation 3' de la levure (Figure 3) comporte quatre éléments généraux. L'EE, avec UAUUAUA comme séquence optimale, est généralement situé 25 à 40 nucléotides en amont du site poly(A) mais peut être plus distant, jusqu'à 75 nucléotides ou plus. Le PE, qui est équivalent aux PAS des métazoaires, est positionné 10-30 nt en amont du site de clivage, avec AAUAAA comme séquence optimale. Deux éléments riches en U sont retrouvés immédiatement avant et après le site de clivage. Tous les éléments constituant le signal poly(A) chez la levure sont apparemment tolérants à des variations significatives. Par exemple, l'EE peut avoir beaucoup de substitutions, particulièrement A → G ou U → C, avec peu ou pas d'effets sur l'efficacité du processing (Guo and Sherman 1996). Il est notable que les transcrits avec une demie vie longue et ceux avec une demie vie courte

semblent avoir des différences systématiques dans l'EE, suggérant des rôles plus complexes pour les signaux poly(A) dans la stabilité des ARNm (Graber 2003). Au-delà des signaux séquence-spécifiques, il a également été suggéré que le contexte structural de l'ARN jouerait un rôle dans la reconnaissance par les complexes poly(A) (Graveley, Fleming et al. 1996), mais peu d'études ont été menées sur ce sujet jusqu'à ce jour.

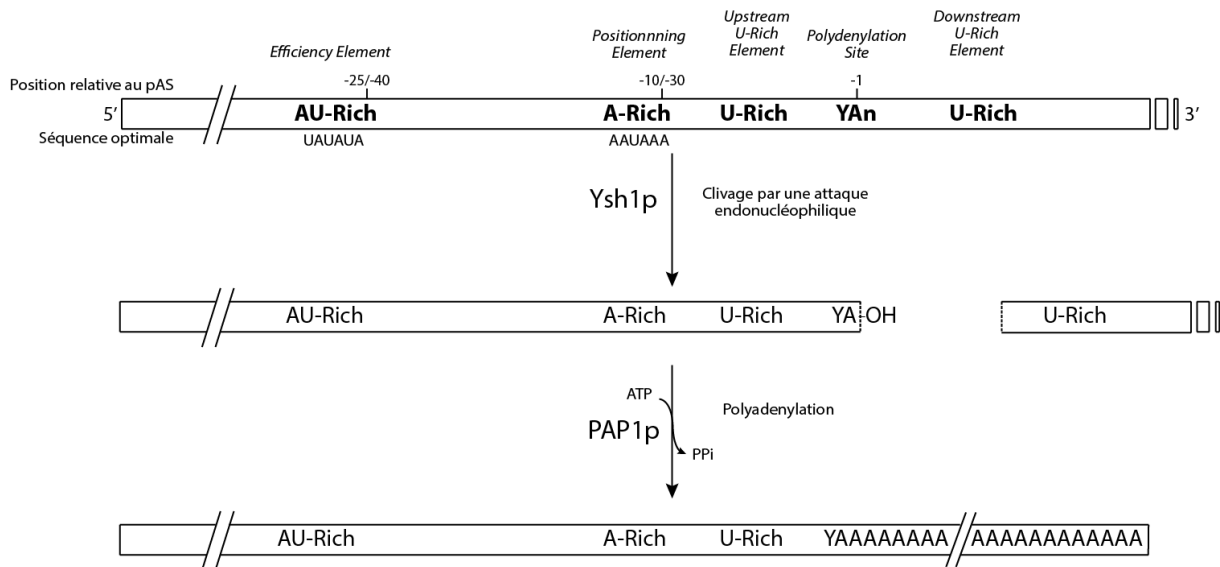


Figure 3 : Description du site de polyadénylation canonique de la levure et des étapes de la maturation 3'.

Chez la levure, les impératifs de séquences déterminant le signal poly(A) sont en tout cas peu astringents. La question de comment la cellule peut reconnaître un site de polyadénylation spécifique de manière précise et fiable dans le contexte génomique reste posée.

1.3.2. Éléments *trans*

La machinerie protéique permettant la maturation 3' des ARN messagers d'eucaryotes contient plusieurs sous-complexes. Chez la levure, ils sont au nombre de trois (Chen and Moore 1992, Kessler, Zhao et al. 1996) : CPF, CF IA et CF IB (cf Figure 4 et Figure 5).

Le facteur de clivage et de polyadénylation (Cleavage and Polyadenylation Factor, CPF), essentiel au clivage et à la polyadénylation, est composé de 20 sous unités, organisées en deux sous-complexes, CF II et PF I (Chen and Moore 1992), environnés de quelques protéines plus ou moins intégrées. Il est retrouvé en interaction avec la Pol II lors des étapes les plus précoces de la transcription, et ne s'en dissocie qu'en aval du site poly(A). Il porte les activités

enzymatiques de clivage (sous unité Ysh1p (Zhelkovsky, Tacahashi et al. 2006, Garas, Dichtl et al. 2008)) et de polyadénylation (sous unité Pap1p) et reconnaît l'ARN au niveau des éléments riches en U qui encadrent le site de clivage (Dichtl and Keller 2001).

Yhh1/Cft1 (153 kDa): Plus grande sous unité du CPF, elle est composée d'une succession de répétitions de motifs WD40 formant 3 β -propeller, identifiés grâce à l'analyse de la séquence car la structure n'a pas été résolue. Ces motifs sont très communs cependant, souvent impliqués dans des interactions protéine-protéine, et une idée de l'architecture de ce domaine peut être donnée grâce à la structure du facteur DDB1 qui partage une homologie de séquence avec Yhh1 (Li, Chen et al. 2006). Fonctionnellement, Yhh1 est indispensable au clivage et à la polyadénylation. Il interagit directement avec le pré-ARNm à proximité du site de clivage, reconnaissant une région riche en A (Dichtl, Blank et al. 2002) via le β -propeller central. Son rôle ne se restreint cependant pas à cela car une interaction avec le domaine C-terminal phosphorylé de la Pol II (CTD) a également été décrite (Dichtl, Blank et al. 2002), ce qui impliquerait Yhh1 dans le couplage de la maturation 3' et de la terminaison de la transcription.

Ydh1/Cft2 (96 kDa): Bien qu'il soit membre de la famille des métallo- β -lactamases, et plus précisément des β -CASP, Ydh1 est dépourvu des résidus impliqués dans la liaison du Zinc, ce qui le rend incapable de catalyser. La résolution de la structure de son domaine β -lactamase confirmera l'absence d'atome de zinc (Mandel, Kaneko et al. 2006). Cependant, Ydh1 reste essentiel à la viabilité cellulaire, et est impliquée dans les deux étapes de clivage et de polyadénylation. Sa fonction précise est toujours inconnue, mais une interaction directe avec le pré-ARNm est suggérée (Zhao, Kessler et al. 1997, Dichtl and Keller 2001) et des interactions directes avec Pcf11 et le CTD de la Pol II ont été mises en évidence (Kyburz, Sadowski et al. 2003). L'homologue de mammifère, CPSF100, est également impliqué dans une hétérodimérisation avec CPSF73, interaction spécifique à la famille des β -CASP et indispensable à l'activité de CPSF73. CPSF73 est également l'endonucléase impliqué dans la maturation des ARNm des histones (Kolev, Yario et al. 2008) et un homologue a également été identifié chez les archées (Dominski, Carpousis et al. 2013), illustrant la très forte conservation de ces mécanismes.

Ysh1/Brr5 (88 kDa): Grâce à des analyses de séquences montrant l'appartenance de Ysh1 à la famille des β -CASP (Callebaut, Moshous et al. 2002) et à l'étude de mutants incapables de liaison au zinc (Ryan, Calvo et al. 2004), cette sous-unité de CPF a toujours été suspectée d'être l'endonucléase responsable du clivage. Mais la preuve directe n'en a finalement été apportée que par la résolution de la structure cristallographique (Mandel, Kaneko et al. 2006). Cependant, la protéine seule n'est pas active et l'homologue de mammifère, CPSF73, nécessite une hétérodimérisation avec CPSF100 (Kolev, Yario et al. 2008), suggérant un mécanisme similaire pour la protéine de levure.

Pta1 (88 kDa): Pta1 est identifié comme un acteur de la maturation 3' depuis la fin des années 90 (Zhao, Hyman et al. 1999). Il semble jouer un rôle architectural majeur en interagissant avec de nombreux autres facteurs tels que Ysh1 (Zhelkovsky, Tacahashi et al. 2006), Ydh1 (Kyburz, Sadowski et al. 2003), Fip1 (Ghazy, He et al. 2009), Pcf11 (Kyburz, Sadowski et al. 2003), Clp1 (Ghazy, He et al. 2009), Pap1 (Ghazy, He et al. 2009) et Ssu72 (He, Khan et al. 2003). La structure du domaine N-terminal de l'homologue de mammifère, la Symplekin, a été résolue, montrant une organisation en HEAT repeat (Kennedy, Frazier et al. 2009) et une interaction avec le CTD de la Pol II et Ssu72 (Xiang, Nagaike et al. 2010).

Ref2 (60 kDa): Identifié très tôt comme un membre de la machinerie poly(A) (Russnak, Nehrke et al. 1995) puis comme une sous-unité de CPF (Nedea, He et al. 2003), Ref2 est un facteur spécifique à la levure qui reste très méconnu. Son rôle dans la maturation des ARNm n'est pas clairement déterminé et nous savons seulement qu'il est requis pour la formation d'une queue polyA normale (Mangus, Smith et al. 2004). Il semble cependant responsable de l'adressage et la spécificité de l'activité de la phosphatase Glc7 au sein de CPF (Nedea, He et al. 2003). Il a également été impliqué dans la maturation des snoRNA (Dheur, Vo le et al. 2003).

Pap1 (65 kDa): Les poly(A) polymérases sont conservées entre les eucaryotes (Martin and Keller 1996), notamment les 500 premiers résidus. La structure de la protéine de levure a été résolue (Bard, Zhelkovsky et al. 2000), de même que celle de bovin (Martin, Keller et al. 2000), révélant une organisation avec trois domaines globulaires. Le site actif est formé par l'interaction entre les domaines C- et N-terminaux. En plus du site actif, un site de liaison à l'ARN a été décrit sur le domaine N-terminal (Raabe, Murthy et al. 1994), mais sa fonction reste peu étudiée. Des interactions avec d'autres facteurs ont par contre été documentées,

notamment entre le domaine N-terminal et Pta1 et Yhh1 (Ezeokonkwo, Ghazy et al. 2012) et entre le domaine C-terminal et Fip1 (Preker, Lingner et al. 1995, Helmling, Zhelkovsky et al. 2001, Meinke, Ezeokonkwo et al. 2008). La poly(A) polymérase de la levure n'est pas indispensable pour le clivage du pré-ARN messager, contrairement à son homologue métazoaire.

Mpe1 (50 kDa): Mpe1 est une sous-unité peu connue de CPF, découverte en 2001 (Vo, Minet et al. 2001). Un doigt de zinc putatif a été décrit par analyse de séquence, mais aucune interaction n'est documentée. *In vitro*, une déplétion de Mpe1 ne perturbe pas le CPF ni n'altère les activités de clivage ou de polyadénylation, mais la maturation ne se fait alors plus de manière spécifique.

Pfs2 (53 kDa): Identifié en 1997 (Keller and Minvielle-Sebastia 1997), Pfs2 ne fut étudié que trois ans plus tard (Ohnacker, Barabino et al. 2000). Cette étude a permis de montrer que Pfs2 est indispensable pour la viabilité cellulaire et la polyadénylation, mais pas pour le clivage. Pfs2 est constituée de répétitions WD40 et interagit avec Rna14, Ysh1, Fip1 (Ohnacker, Barabino et al. 2000) et Clp1 (Ghazy, Gordon et al. 2012) et une étude chez *S.pombe* montre des défauts de terminaison pour des mutants de Pfs2 (Wang, Asakawa et al. 2005). L'homologue de mammifère, WDR33, a très récemment été montré comme étant la sous-unité de CPF avec la meilleure spécificité pour l'hexamère AAUAAA (Elmar Wahle, 2014 RNA Society Meeting, données non publiées).

Pti1 (47 kDa): Composant essentiel du CPF (Skaar and Greenleaf 2002), impliqué dans la maturation des snoRNA et des mRNA (Dheur, Vo le et al. 2003), Pti1 partage une forte homologie de séquence avec Rna15. Sa fonction n'est pas clairement établie, mais il a été proposé qu'il œuvre au découplage du clivage et de la polyadénylation lors de la maturation des transcrit non polyadénylés, possiblement via la formation d'un CF IA alternatif (Dheur, Vo le et al. 2003).

Fip1 (36 kDa): Fip1 a très tôt été identifié comme un facteur interagissant avec Pap1 (Preker, Lingner et al. 1995), et cette interaction essentielle (Meinke, Ezeokonkwo et al. 2008) régule l'activité de la polymérase (Helmling, Zhelkovsky et al. 2001). Bien que Fip1 soit déstructurée en solution, son rôle semble de lier la polymérase à CPF, via un linker permettant l'ancrage de protéines tierces (Ezeokonkwo, Zhelkovsky et al. 2011).

Swd2 (37 kDa): Swd2 a été décrite bien avant d'être associée au complexe CPF. Composée de sept répétitions WD40, Swd2 est en effet décrite comme une sous-unité du complexe COMPASS impliqué dans la méthylation de la lysine 4 sur l'histone H (Cheng, He et al. 2004). Une déplétion de Swd2 n'affecte pas la maturation 3' de précurseurs *in vitro*, mais des mutants de Swd2 entraînent un défaut de terminaison sur quelques gènes transcrits par la Pol II *in vivo*.

Glc7 (36 kDa): Glc7 est une serine-thréonine phosphatase impliquée dans de nombreux processus biologiques car sa spécificité est faible (Lenssen, James et al. 2005) et elle doit être couplée à une sous-unité régulatrice. Dans le cadre de la maturation 3', Ref2 semble être la sous-unité responsable de l'association de Glc7 avec CPF (Nedea, He et al. 2003). Plusieurs hypothèses quant à son substrat et son rôle lors de la polyadénylation ont été proposées (Garcia-Gimeno, Munoz et al. 2003, He and Moore 2005), mais ce n'est que récemment que la phosphorylation de la tyrosine 1 (Tyr1-P) du CTD de la Pol II a été clairement identifiée (Schrieck, Easter et al. 2014). Glc7 a donc un rôle central dans la coordination de la transcription et la maturation 3' (cf 1.7.1).

Yth1 (25 kDa): Yth1, une des plus petites sous-unités du CPF, est composée de cinq doigts de zinc impliqués dans des interactions avec l'ARN ou avec d'autres facteurs protéiques. *In vitro*, Yth1 montre une sélectivité pour la liaison aux ARN riches en U (Barabino, Hubner et al. 1997), comme on en trouve de part et d'autre du site polyA. La partie C-terminale de la protéine est spécifiquement requise pour lier Fip1 et la Pap1 au CPF, impliquant Yth1 dans la reconnaissance du site poly(A) et dans la constitution d'un CPF compétent.

Ssu72 (23 kDa): Phosphatase associée à TFIIIB, Ssu72 a d'abord été associée à la sélection du site d'initiation de la transcription (Sun and Hampsey 1996). Elle a ensuite été identifiée comme membre du CPF, nécessaire à l'activité de clivage (Dichtl, Blank et al. 2002, He, Khan et al. 2003). Son substrat est le CTD phosphorylé en Sérine 5 (Krishnamurthy, He et al. 2004), dont le processing est requis pour la maturation 5' du pré-ARNm. La structure de l'homologue humain en complexe avec un heptapeptide du CTD et la région d'interaction de la Symplekine a été résolue (Xiang, Manley et al. 2012).

Syc1 (20,8 kDa): Impliqué dans la maturation des ARN polyadénylés et non polyadénylés, Syc1 est un paralogue de la région C-terminale de Ysh1. Son rôle est peu exploré, mais il semble

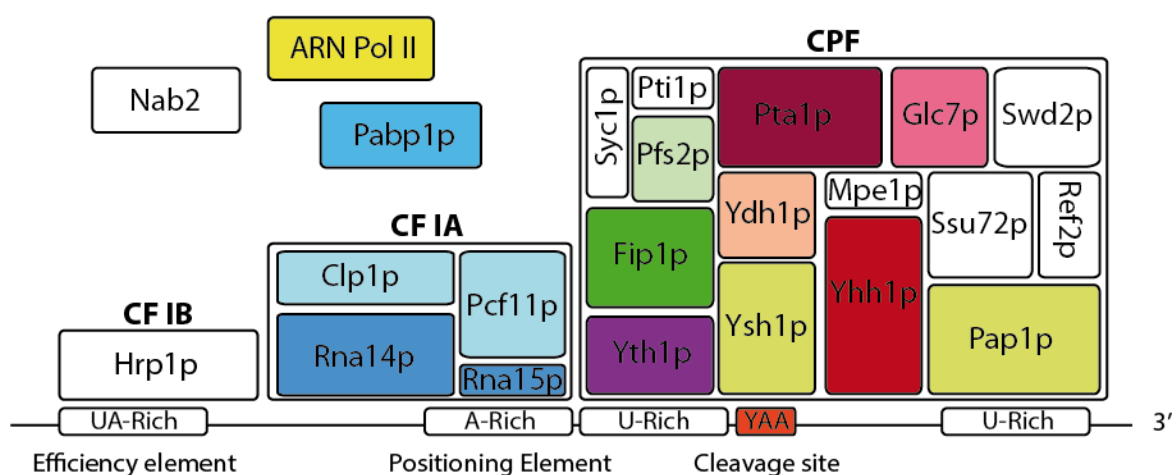
être un régulateur négatif de la maturation 3', peut-être en mimant l'extrémité essentielle de l'endonucléase Ysh1 (Zhelkovsky, Tacahashi et al. 2006).

Le facteur de clivage IB (Cleavage Factor IB, CF IB) est composé d'une seule sous unité, Hrp1p, qui n'a pas d'homologue dans le système mammifère, et qui est impliquée dans la reconnaissance spécifique de l'Efficiency Element. Il est indispensable au clivage et à la polyadénylation. C'est une protéine organisée en trois domaines, dont le domaine central est constitué de deux RRM dont la structure a été résolue en complexe avec le EE (Perez-Canadillas 2006, Leeper, Qu et al. 2010). Il a été montré que l'interaction de Rna15p et de Hrp1p avec Rna14p augmentait l'efficacité de la reconnaissance de la séquence d'ARN (Gross and Moore 2001, Noble, Walker et al. 2004).

Nab2 n'est pas copurifié avec la machinerie de polyadénylation, mais est impliqué dans l'étape de synthèse de la queue polyA en contrôlant la longueur de la queue polymérisée (Viphakone, Voisinet-Hakil et al. 2008).

Il existe également quelques facteurs indépendants : PABP1 et l'ARN polymérase II sont les principaux. PABP1 (Poly(A) Binding Protein 1) interagit avec la queue polyA et permet le contrôle de sa longueur (Amrani, Minet et al. 1997).

A. *S. cerevisiae*



B. Mammifères

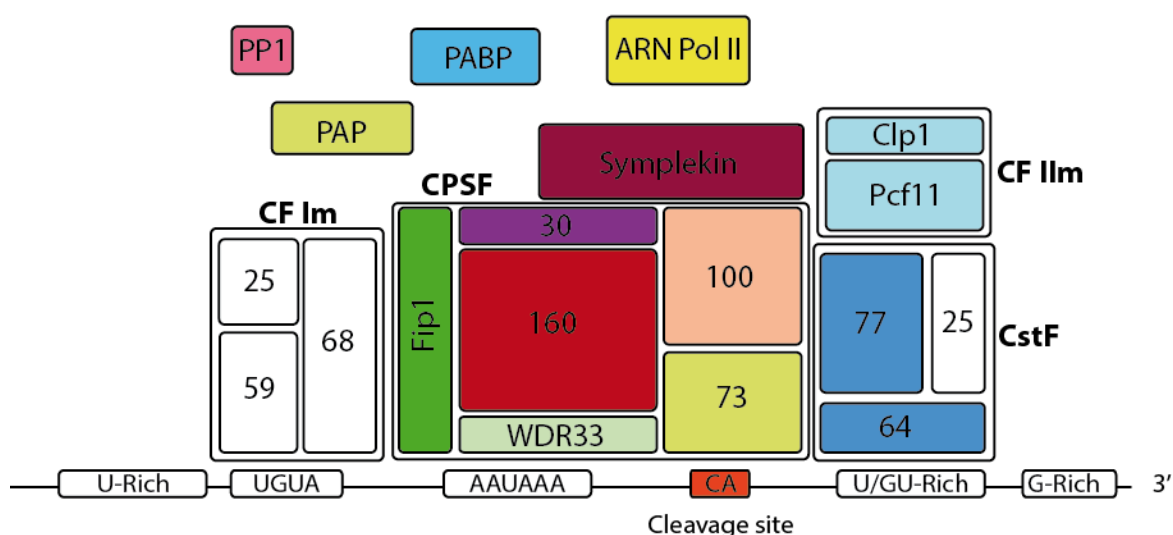


Figure 4 : Schéma récapitulatifs des machineries de maturation 3' chez *Saccharomyces cerevisiae* (A) et chez les mammifères (B). Le jeu de couleur met en avant les homologues conservés entre les deux systèmes. Les facteurs indiqués en blanc n'ont pas d'homologues chez les mammifères (ou la levure) ou leurs homologues ne sont pas impliqués dans la maturation 3'. La représentation met en avant les principales interactions ARN-protéine et ne cherche pas à être exact sur les interactions protéines-protéines et la stœchiométrie des facteurs.

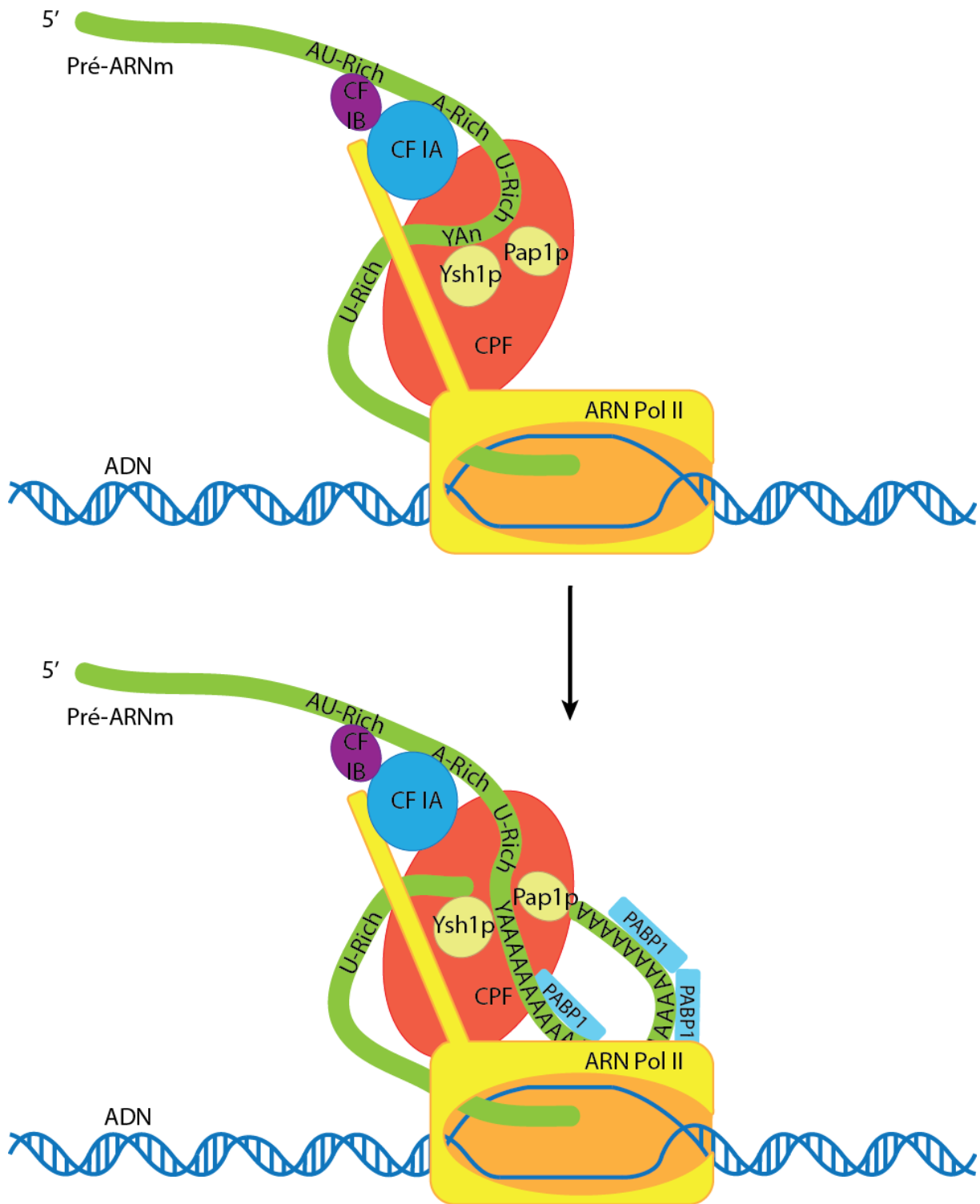


Figure 5 : Schéma représentant les principaux complexes de la machinerie de maturation 3' des ARNm chez la levure, en complexe avec le pré-ARNm naissant sur lequel sont indiqués les motifs définissant le signal poly(A).

1.3.3. Focus sur CF IA

Le facteur de clivage IA (Cleavage Factor IA, CF IA) est constitué de quatre sous-unités, Rna14p, Rna15p (Minvielle-Sebastia, Winsor et al. 1991, Minvielle-Sebastia, Preker et al.

1994), Pcf11p (Amrani, Minet et al. 1997) et Clp1p (Minvielle-Sebastia, Preker et al. 1997). Il a tout d'abord été identifié en compagnie de CF IB, en tant que fraction chromatographique indispensable au clivage et à la polyadénylation (Chen and Moore 1992). Le *distinguo* entre CF IA et CF IB fut fait quelques années plus tard, grâce à l'amélioration de la purification (Kessler, Zhao et al. 1996). L'identité des facteurs composant CF IA n'était alors pas complètement claire, mais son rôle indispensable à la fois pour le clivage et la polyadénylation fut confirmé.

Une étude de mutants thermosensibles attira l'attention sur les gènes RNA14 et RNA15 bien avant que leur produit ne soit caractérisé. Les phénotypes induits par les mutations dans les deux gènes se caractérisaient par une réduction de la stabilité des ARNm et de la longueur de la queue polyA (Minvielle-Sebastia, Winsor et al. 1991) et ces deux protéines ne furent pas identifiées lors des toutes premières purifications de la machinerie 3' (Chen and Moore 1992). De cette première étude, appuyée par une localisation nucléaire pour les deux protéines (Bonneaud, Minvielle-Sebastia et al. 1994), il fut proposé que Rna14 et Rna15 interagissent ensemble. Finalement, l'hypothèse de leur implication dans la stabilité des ARNm fut révoquée au profit d'une implication dans la polyadénylation lorsqu'il fut mis en évidence que l'activité CF I de la maturation 3' (Chen and Moore 1992) complétait les mutants (Minvielle-Sebastia, Preker et al. 1994). Une étude génétique confirma cette hypothèse (Mandart and Parker 1995), proposant même que Rna14 et Rna15 soient impliquées dans le choix du site de polyA. Cette théorie prit encore plus de poids lorsque les deux protéines furent identifiées comme composants du CF IA (Kessler, Zhao et al. 1996).

Rna14p (Figure 6) est la plus grande protéine de CF IA, un polypeptide essentiel de 677 résidus pour un poids moléculaire de 80 kDa. Il présente une forte homologie de séquence avec son homologue mammifère, CstF77 (Takagaki and Manley 1994), et sont toutes deux des protéines fortement structurées, dont environ 80% de la séquence est organisée en 12 répétitions de domaine Half-a-TPR (HAT) (Bai, Auperin et al. 2007, Legrand, Pinaud et al. 2007).

Le tétratricopeptide (TPR) est une séquence de 34 acides aminés, plus ou moins dégénérée, fréquemment retrouvée en répétition de dimères et connue pour permettre l'organisation de complexes multiprotéiques. Le domaine HAT est un variant du motif TPR qui

possède une structure conservée en hélice-boucle-hélice où les deux hélices forment une association antiparallèle stable.

Ce « corps » de la Rna14p engage une homodimérisation (Bai, Auperin et al. 2007, Legrand, Pinaud et al. 2007) essentielle pour la viabilité cellulaire et constituant une surface importante, conservée par endroit. L'extrémité C-terminale de la protéine est impliquée dans une hétérorimérisation avec Rna15p. Cette interaction est elle aussi essentielle et conservée. Rna14p interagit également avec d'autres facteurs de maturation 3', notamment Hrp1p ou le CTD de la Pol II.

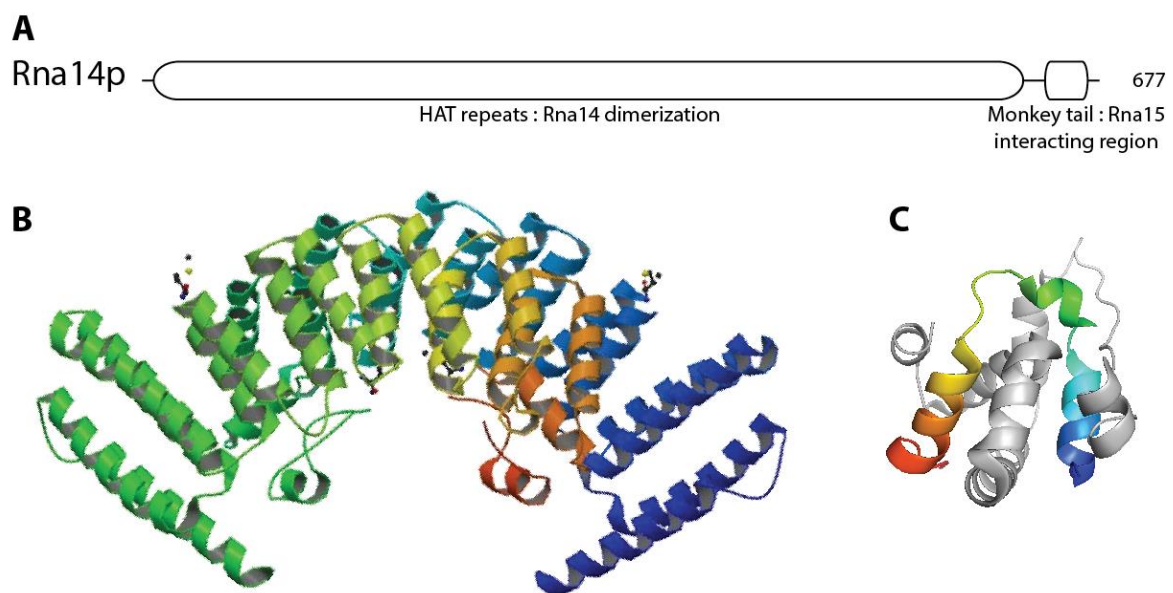


Figure 6: Domaines connus de Rna14. **A.** Schéma de la séquence de Rna14p. Les domaines dont la structure est résolue chez la levure ou un homologue sont représentés en ovale. **B.** Hétérodimère des domaines HAT de CstF-77 (Bai, Auperin et al. 2007, Legrand, Pinaud et al. 2007) **C.** Domaine Monkey-tail de Rna14p (couleur) en complexe avec le domaine Hinge de Rna15p (blanc) (Moreno-Morcillo, Minvielle-Sebastia et al. 2011)

Rna15p (Figure 7) est le plus petit composant de CF IA, mais il a très tôt attiré l'attention, car une analyse de sa séquence révélât un domaine pouvant potentiellement lier l'ARN, ce qui expliquerait que CF IA soit un facteur indispensable au clivage et à la polyadénylation. Trois domaines conservés sont en fait détectés, et la structure de ces trois domaines isolés a été résolue. En N-terminal se trouve le RRM responsable de la reconnaissance de l'élément PE riche en A *in vivo* (Gross and Moore 2001), bien que son affinité *in vitro* soit plus importante pour des séquences riches en U (Takagaki and Manley 1997). Cette sélectivité serait apportée par l'interaction avec Hrp1 et Rna14 qui modifierait le

site reconnaissant l'ARN dans le contexte CF I (Gross and Moore 2001, Gross and Moore 2001, Leeper, Qu et al. 2010). Le domaine central de la protéine est responsable de l'interaction très étroite observée entre Rna14p et Rna15p (Moreno-Morcillo, Minvielle-Sebastia et al. 2011). Les premiers mutants décrits dans la littérature sont affectés dans cette région. La région C-terminale de Rna15p est un ensemble d'hélices α repliées entre elles pour former une surface d'interaction constituée de résidus conservés, permettant le contact avec Pcf11p (Qu, Perez-Canadillas et al. 2007).

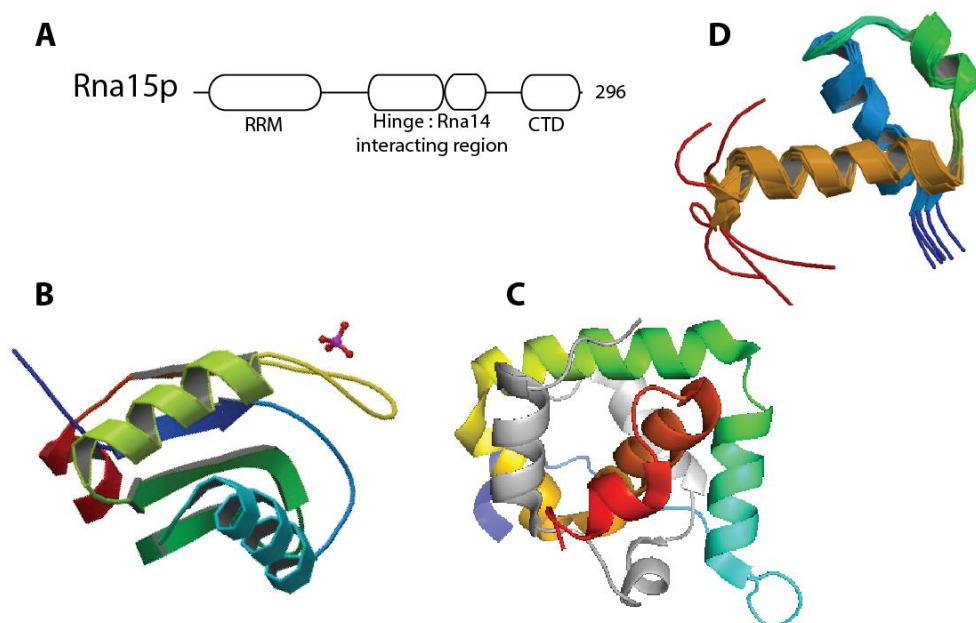


Figure 7 : Domaines connus de Rna15. **A.** Schéma de la séquence de Rna15p. Les domaines dont la structure est résolue chez la levure ou un homologue sont représentés en ovale. **B.** Structure du RRM de Rna15p (Pancevac, Goldstone et al. 2010) **C.** Domaine Hinge de Rna15p (couleur) en complexe avec le domaine Monkey-tail de Rna14p (blanc) (Moreno-Morcillo, Minvielle-Sebastia et al. 2011) **D.** Domaine C-terminal de CstF-64 (Qu, Perez-Canadillas et al. 2007)

Pcf11p (Figure 8) est une protéine essentielle chez la levure, requise pour le clivage mais pas pour la polyadénylation (Amrani, Minet et al. 1997). Les principaux partenaires décrits pour Pcf11p sont Rna15p, Clp1p et le Ser2-P CTD de la Pol II. De fait, cela amena à décrire deux fonctions complémentaires pour cette protéine : un rôle dans la terminaison de la transcription et un rôle dans la maturation 3' (Sadowski, Dichtl et al. 2003). Le N-terminal de Pcf11p interagit *in vitro* avec le CTD de l'ARN polymérase II phosphorylé au niveau de la sérine 2 mais déphosphorylé sur la tyrosine 1, état caractéristique du complexe transcriptionnel arrivé au site de polyadénylation (Noble, Hollingworth et al. 2005, Mayer, Heidemann et al. 2012, Gu, Eick et al. 2013). Cette interaction est essentielle à la terminaison

de la transcription (Zhang, Fu et al. 2005, Hollingworth, Noble et al. 2006), et le rôle de Pcf11p dans ce processus biologique est détaillé plus en détail dans le chapitre correspondant. Par contre, la moitié C-terminale de la protéine est impliquée dans des interactions d'une part avec Rna15p et d'autre part avec Cpl1p, faisant ainsi de Pcf11p le nœud central du complexe CF IA. L'interaction avec Rna15p a été caractérisée structuralement (Qu, Perez-Canadillas et al. 2007). L'interaction avec Cpl1p est également bien connue, notamment grâce à la résolution de la structure de Cpl1p en complexe avec un peptide partenaire (Noble, Beuth et al. 2007), mais aussi via l'étude de mutants. Ces deux interactions sont essentielles à la maturation 3'. Pcf11p interagit également avec les sous unités Yhh1p/Cft1, Ydh1p/Cft2, Ysh1p/Brr5p et Pta1p du CPF. Cette interaction avec l'endonucléase et ses cofacteurs est probablement impliquée dans le fait que Pcf11p/Cpl1p ne sont requis que pour l'activité de clivage et non celle de polyadénylation. Trois autres éléments méritent d'être notés chez Pcf11p. Tout d'abord l'existence en N-terminal d'un second domaine, juste après le domaine impliqué dans la terminaison, dont le rôle est inconnu, mais qui semble conservé (résultats du laboratoire non publiés). Conservés également, on retrouve deux motifs putatifs de liaison au zinc, décrits d'après la séquence uniquement, de part et d'autre de la zone d'interaction avec Clp1p (ce travail). Le dernier élément est spécifique à la levure *S. cerevisiae*, et dans une certaine mesure à *K. lactis*. Il s'agit d'une succession de 20 glutamines, à la fonction inconnue, bien que l'on puisse supposer qu'elles introduisent une forte souplesse dans la protéine.

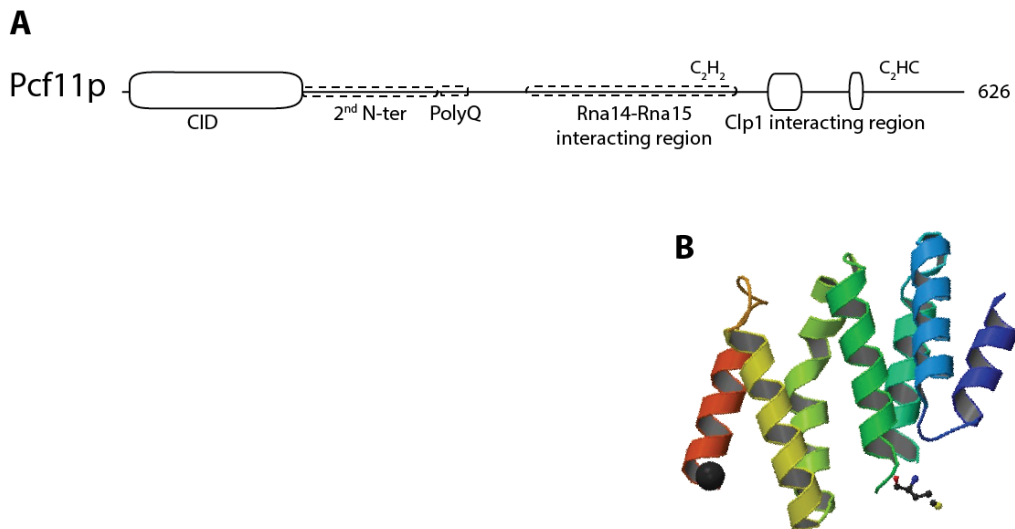


Figure 8 : Domaines connus de Pcf11. **A.** Schéma de la séquence de Pcf11p. Les domaines dont la structure est résolue chez la levure ou un homologue sont représentés en ovale. **B.** Domaine N-terminal de Pcf11 (Meinhart and Cramer 2004)

Clp1p (Figure 9) fut la dernière protéine du CF IA à être identifiée comme telle. C'est une protéine capable de lier l'ATP grâce à un motif de type Walker A conservé, mais aucune hydrolyse n'a pu être mise en évidence chez la levure. Et chez les homologues d'archée et de métazoaires pour lesquels une hydrolyse de l'ATP est détecté, cela ne se fait pas dans le cadre de la maturation 3' des ARNm, mais dans la maturation des ARNt (Dikfidan, Loll et al. 2014, Karaca, Weitzer et al. 2014, Schaffer, Eggens et al. 2014). Pourtant, chez la levure, des mutants perturbant la liaison à l'ATP sont délétères pour la maturation 3' (Ghazy, Gordon et al. 2011, Haddad, Maurice et al. 2011). Au sein de CF IA, Clp1p interagit exclusivement avec Pcf11p mais il interagit également avec l'endonucléase Ysh1p et des sous-unités du CPF (Holbein, Scola et al. 2011). Cette interaction n'a pas été étudiée de manière approfondie, mais peut se synergiser avec celles engagées par Pcf11p et expliquer pourquoi Clp1p et Pcf11p sont indispensables uniquement pour l'étape de clivage en participant au recrutement/positionnement de l'enzyme.

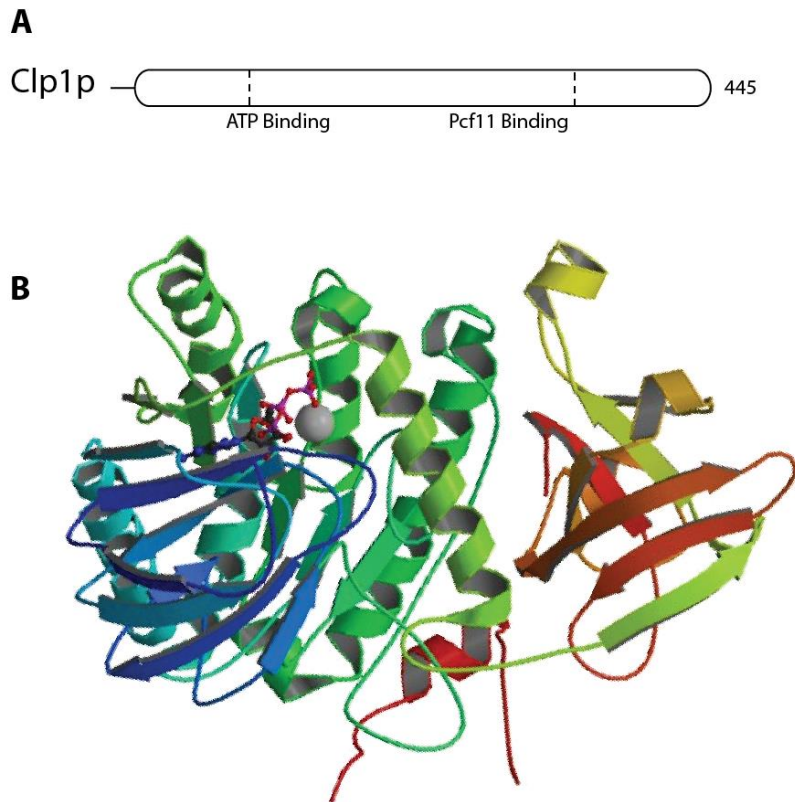


Figure 9 : Domaines connus de Clp1. **A.** Schéma de la séquence de Clp1p. Les domaines dont la structure est résolue chez la levure ou un homologue sont représentés en ovale. **B.** Structure de Clp1p en complexe avec une molécule d'ATP et un peptide de Pcf11p (non visible) (Noble, Beuth et al. 2007)

De manière plus anecdotique, il est également proposé que le complexe CF IA joue un rôle dans la réinitiation de la transcription (Al Husini, Kudla et al. 2013).

1.4. Comparaison machinerie Levure / Mammifères

Dans le système métazoaire, le signal poly(A) est bien mieux conservé. L'exemple le plus évident est l'hexamère AAUAAA situé 10 à 30 nucléotides en amont du site de clivage (Colgan and Manley 1997), bien que plus de 10 variants sous-optimaux aient également été identifiés (e.g. AUUAAA) (Tian, Hu et al. 2005). Environ 20 à 40 nucléotides en aval du site de clivage se trouve une séquence moins conservée, riche en U ou en GU nommée Downstream sequence élément (DSE). La variation de séquence possible au niveau du DSE est proposée comme étant un mécanisme compensatoire à l'utilisation d'hexamères sous-optimaux (Zarudnaya, Kolomiets et al. 2003, Nunes, Li et al. 2010). Entre l'hexamère et le DSE se trouve le site de clivage, où la réaction endonucléotidique à préférentiellement lieu après un dinucléotide CA (Chen, MacDonald et al. 1995), bien que cela ne soit pas un impératif. L'efficacité du 3'

processing est régulée par des séquences régulatrices additionnelles (Legendre and Gautheret 2003, Hu, Lutz et al. 2005). Par exemple, un élément riche en U (U-Rich sequence element, USE) situé en amont et un élément UGUA stimulent la réaction (Gilmartin, Fleming et al. 1995, Moreira, Takagaki et al. 1998, Yang, Gilmartin et al. 2010). De plus, une étude plus récente a rapporté que certains signaux poly(A) humains ne nécessiteraient qu'un DSE et une séquence riche en A en amont (Nunes, Li et al. 2010).

Les protéines du coeur requises pour la réaction de clivage/polyadénylation sont par contre conservées entre les métazoaires et la levure, alors même que les signaux poly(A) diffèrent (Proudfoot 2004, Mandel, Bai et al. 2008, Millevoi and Vagner 2010), mais elles sont organisées différemment. Certains facteurs conservent leur capacité de liaison à l'ARN, comme Ysh1p (levure) et CPSF-73 (mammifère), qui sont tous deux les endonucléases. De la même manière, Yth1p (levure) et CPSF-30 (mammifère) reconnaissent tout deux des séquences riches en U. *A contrario*, on observe de fortes différences entre certains homologues. Yth1p (levure) est impliqué dans la reconnaissance de l'ARN, avec une spécificité pour les séquences riches en A à proximité de site poly(A), alors que son homologue CPSF-160 interagit extrêmement spécifiquement avec l'hexamère AAUAAA très fortement conservé. De très récents travaux de Lars Schönemann (données non publiées) viennent cependant nuancer ceci car ils mettent en évidence que la spécificité pour l'hexamère AAUAAA serait apportée par WDR33. Si cela est avéré, les fonctions de CPSF-160 et Yth1p seraient alors légèrement différentes. De la même manière, hFip1 reconnaît des séquences riches en U grâce à un domaine C-terminal absent chez la levure. Rna15p est également un cas intéressant. *In vitro*, il montre une forte affinité pour les séquences riches en U, ce qui ferait écho à la reconnaissance spécifique du DSE riche en U par son homologue métazoaire CstF-64. Mais *in vitro*, dans le contexte CF IA, Rna15p montre une spécificité pour la séquence riche en A du PE. Cela semble être induit par son interaction avec Rna14p et Hrp1p, qui n'a pas d'homologue mammifère et qui lie très spécifiquement le EE riche en AU (Leeper, Qu et al. 2010, Pancevac, Goldstone et al. 2010).

1.5. Les autres maturations de l'extrémité 3'

Chez les eucaryotes, une seule famille d'ARN messagers est connue pour ne pas comporter de queue poly(A) une fois mature, ce sont les ARNm des histones. A la place, le messenger se

termine avec une structure tige-boucle extrêmement conservée. Cette caractéristique unique est formée par un complexe riboprotéique spécifique qui reconnaît deux caractéristiques principales sur la séquence : d'une part la tige-boucle, et d'autre part une séquence riche en purine, en aval du segment structuré (Dominski and Marzluff 1999). Le clivage a lieu entre ces deux éléments, catalysé par Ysh1p, la même endonucléase donc que dans la maturation classique. Par contre, le clivage de l'ARNm d'histone n'est pas suivi de polyadénylation. La machinerie de maturation est moins importante, constituée de seulement 7 protéines et d'un snRNA (Yang, Sabath et al. 2012, Tan, Marzluff et al. 2013). Le snRNA est responsable de la reconnaissance de la séquence riche en purine tandis qu'un facteur protéique lie le motif tige-boucle. Trois facteurs seulement sont partagés entre la machinerie classique et celle de maturation des histones : l'endonucléase Ysh1p, son partenaire d'hétérodimérisation Ydh1p et la symplektine/Pta1p.

Mais il n'existe pas que les ARN messagers dans la cellule, ceux-ci représentent même une proportion plutôt minoritaire. L'activité ARN polymérase II par exemple, ne s'exerce par uniquement sur des gènes codant pour des protéines car elle est également responsable de la synthèse de petits ARN non codants, notamment la majorité des petits ARNs nucléaires. Or ces transcrits ne sont pas polyadénylés. La machinerie de maturation, dont le recrutement est liée à la Ser7-P du CTD, est spécifique est encore peu étudié. Certaines sous-unités semblent en tout cas similaires à celle de la maturation classique, sans pour autant que cela soit les mêmes.

En ce qui concerne les longs ARN non-codants, les quelques cas connus sont maturés par la RNase P, l'enzyme responsable de la maturation 5' des ARNs (Zhang and Zhu 2014). Ce clivage est effectué suite à la reconnaissance d'une structure particulière qui est retrouvée en aval des longs ARN non codants étudiés. Et en cherchant dans le génome pour cette structure, deux gènes non codants particuliers ont été découverts : MALAT1 et MEN β . Ces deux séquences ont en effet la particularité de présenter un signal poly(A) classique mais également un signal structural en amont entraînant pour la large majorité de ces transcrits une maturation par la RNase P. Quelle que soit la voie de maturation engagée, ces transcrits resteront dans le noyau, indiquant que la maturation 3' n'est pas responsable de leur localisation subcellulaire.

La maturation des ARN télomérique est également intéressante (Wilusz and Spector 2010, Peart, Sataluri et al. 2013). Il y a en effet un site poly(A) à l'extrémité 3' de l'ARN des télomérases, et pourtant 95% des transcrits se terminent en amont et ne sont pas polyadénylés. En fait, les formes les plus longues sont même inactives *in vivo*. Ici, pas de ressemblance avec des ARNt, mais une forte similarité avec la bordure 5' d'un site d'épissage. Et en effet, c'est le spliceosome qui a été montré comme responsable du clivage de ces ARN, bien que sans que le mécanisme ne soit pas connu, il n'enchaîne pas avec la ligation de l'extrémité 5' sur l'extrémité 3' qu'il génère.

1.6. Polyadénylation alternative et sélection du site poly(A)

Comme nous avons pu le voir, les impératifs de séquence pour la définition du site de polyadénylation sont relativement lâches. Ce n'est donc pas une surprise que de nombreux signaux poly(A) aient été identifiés dans des endroits variés de la région transcrite. Les signaux poly(A) putatifs sont fréquemment retrouvés dans les régions riches en A-U telles que les introns (Tian, Pan et al. 2007), les régions non traduites (Lee, Ji et al. 2008) et les régions intergéniques (Lopez, Granjeaud et al. 2006). De plus, des variants de l'hexamère AAUAAA peuvent également être fonctionnels chez les mammifères. Des études bioinformatiques ont montrées que le motif AAUAAA représente seulement 53% des sites chez l'humain et 59% chez la souris. La séquence AUUAAA est retrouvé sur environ 16% des signaux poly(A) (Tian, Hu et al. 2005). Les autres hexamères détectés sont UAUAAA, AGUAAA, AAGAAA, AAUAUA, AAUACA, CAUAAA, GAUAAA, AAUGAA, UUUAAA, ACUAAA et AAUAGA (Beaudoing, Freier et al. 2000, Tian, Hu et al. 2005). De manière notable, environ 54% des gènes humains ont un signal poly(A) alternatif, et on compte 2,1 site de polyadénylation utilisé pour chaque gène (Tian, Hu et al. 2005). Les polyadénylations introniques sont relativement fréquentes et retrouvées dans environ 20% des gènes humains (Tian, Pan et al. 2007). Des 3'UTR extrêmement longs, plus de 10kb, ont également été détectés expérimentalement dans des cellules de mammifères (Moucadel, Lopez et al. 2007) tandis que un petit nombre d'évènement de polyadénylation dans le 5'UTR ont également été listés, bien que leur efficacité et leur fonction soient méconnues (Tian, Pan et al. 2007). De plus, le nombre de transcrits avec des sites de polyadénylation alternative peut être sous-estimé car toutes les analyses de bio-informatique sont effectuées sur des bases de données d'EST, qui ne

représenteraient que 10% des transcrits, avec une absence d'information sur la transcription spécifique à certains tissus ou stades de développement (Moucadel, Lopez et al. 2007). En effet, des données plus récentes de « deep-sequencing » ont révélées que la polyadénylation alternative était deux à trois fois plus fréquente si l'on comparait deux tissus que si l'on compare deux individus (Wang, Sandberg et al. 2008). De plus, les pré requis de séquence pour le signal poly(A) semblent moins conservés et les signaux non canoniques seraient bien plus communs que ce qui était précédemment démontré (Nunes, Li et al. 2010).

La polyadénylation alternative influence de nombreux aspects de l'expression génique (Edwalds-Gilbert, Veraldi et al. 1997). Des signaux de poly(A) placés en tandem dans le même exon terminal vont produire des transcrits avec la même région codante mais différents 3'UTRs. Cela peut conduire à des variations de stabilité, d'efficacité de traduction ou de localisation sub-cellulaire. La combinaison de la polyadénylation alternative avec un épissage alternatif est bien documentée avec le cas d'école des chaînes lourdes d'immunoglobulines. L'utilisation du site proximal conduirait à la production d'une forme plus courte sécrétée tandis que le site distal permet la synthèse des récepteurs membranaires. Durant l'activation des cellules B, il y a un basculement dans l'utilisation des sites en passant d'un usage préférentiel du site distal à un usage du site proximal. Il a été montré qu'un haut niveau d'expression de CstF-64 est à l'origine de cette maturation alternative.

D'une manière générale, des études de génomique suggèrent que les cellules en prolifération ont tendance à favoriser l'utilisation des sites proximaux pour maximiser l'expression génique. En accord avec ces observations, les cellules cancéreuses montrent des 3'UTRs particulièrement courts bien que les mécanismes de sélection ne soient pas clairs.

1.7. Couplage avec les autres mécanismes

1.7.1. Transcription & Terminaison

La dernière étape du cycle de transcription implique le relargage de l'ARN polymérase II de la matrice d'ADN. Des études menées durant la fin des années 80 démontrèrent que la terminaison de la transcription est étroitement couplée à la maturation 3' du transcrit (Whitelaw and Proudfoot 1986, Logan, Falck-Pedersen et al. 1987, Connelly and Manley 1988). Plus précisément, les signaux génétiques impliqués dans le clivage et la polyadénylation furent

identifiés comme indispensables à une terminaison correcte. Alors que les sites de pause de la Pol II promeuvent la terminaison (Yonaha and Proudfoot 1999, Gromak, West et al. 2006), il semble ne pas y avoir de « séquence de terminaison » consensus et le relargage de la polymérase à lieu en aval du site poly(A), de manière aléatoire, et jusqu'à des distances de plus de 1kb (Tran, Kim et al. 2001, Orozco, Kim et al. 2002).

La notion selon laquelle la maturation 3' serait couplée à la transcription *in vivo* fut initialement appuyée par des mutants de l'ARN polymérase II résistants à l' α -amanitine et délétés du C-terminal Domain. Ces mutants étaient en effet défectueux pour la polyadénylation et d'autres étapes de la maturation, comme l'épissage et l'ajout de la coiffe (McCracken, Fong et al. 1997). Des essais *in vitro* amenèrent ensuite les preuves que la Pol II, ou le CTD recombinant, est requis pour le clivage au site poly(A), et que le clivage est transcription dépendant (Hirose and Manley 1998). L'interaction entre la machinerie de maturation 3' et la Pol II se fait de manière précoce dans la transcription, comme le montre la purification des composants de CPSF au cours d'une immunoprécipitation de TFIID, et ensuite la démonstration que CPSF est transféré sur la polymérase II durant l'initiation de la transcription, au promoteur (Dantonel, Murthy et al. 1997). Ces observations ont plus tard été corroborées par des analyses CHIP du positionnement de la machinerie de polyadénylation le long des gènes (Calvo and Manley 2005).

Deux modèles dominants sur le couplage entre la maturation 3' et la terminaison sont à ce jour proposés (Buratowski 2005). L'un, formulé comme le modèle allostérique ou anti-terminateur, propose que le basculement de l'élongation à la terminaison implique un changement dans les facteurs associés à la Pol II qui est induit lors de la rencontre avec les signaux poly(A) sur l'ARN naissant. En accord avec ce modèle, Tran et al. (Tran, Kim et al. 2001) ont utilisés une inhibition cis-antisens pour démontrer que la reconnaissance des signaux poly(A) au niveau de l'ARN est nécessaire à la terminaison, et que la terminaison peut avoir lieu en absence de clivage. D'autres approches en CHIP ont montrées la présence de certains facteurs d'élongation au-delà des régions transcrites, mais pas après le site poly(A) (Ahn, Kim et al. 2004), ce qui serait en accord avec un réarrangement des facteurs sur le CTD. Le second modèle, connu comme modèle torpille, suggère qu'une exonucléase cible l'extrémité 5' non protégée libérée par la réaction de clivage et dégrade l'ARN jusqu'à ce qu'elle rencontre et désengage la polymérase. Ce modèle est quant à lui appuyé par l'observation de défaut de la

terminaison dans des levures dépourvues de la nucléase Rat1 (Kim, Krogan et al. 2004) ou dans des cellules HeLa chez lesquelles l'expression de l'orthologue de Rat1 Xrn2 est supprimé par ARNi (West, Gromak et al. 2004). Comme cela est le cas pour beaucoup de processus biologiques, les deux modèles coexistent peut être *in vivo* et leur efficacité relative varie selon les gènes (Kim, Vasiljeva et al. 2006, Luo, Johnson et al. 2006).

La réaction de maturation 3' est imbriquée avec le complexe de transcription et la structure de la Pol II joue un rôle majeur dans cette intégration. L'holoenzyme Pol II est un complexe de 550kDa en 12 protéines chez la levure (Cramer, Bushnell et al. 2000, Cramer, Bushnell et al. 2001). Sous le tunnel de sortie de l'ARN se trouve le domaine C-terminal (C-Terminal Domain, CTD), qui est un domaine relativement déstructuré semblant s'éloigner du corps de la pol II (Cramer, Bushnell et al. 2000, Gnatt, Cramer et al. 2001). Il contient 26 répétitions de l'hexamère YSPTSPS chez la levure, 52 chez les mammifères (Zhang, Kim et al. 2006, Moore and Proudfoot 2009). Les résidus sérine du CTD sont l'objet de phosphorylations réversibles durant le cycle de transcription et interagissent directement avec des composants des machineries de maturation du pré-ARNm (Fong and Bentley 2001). La phosphorylation de certains résidus corrèle spécifiquement avec certaines étapes du cycle transcriptionnel. La phosphorylation de la sérine 5 (Ser5P) est abondante dans les étapes précoces de l'élongation (les premières centaines de nucléotides) puis diminue rapidement (Komarnitsky, Cho et al. 2000). La sérine 2 phosphorylée (Ser2P) est étroitement associée aux étapes suivantes de l'élongation, plus productives, durant lesquelles l'épissage et la maturation 3' surviennent. Ser5 et Ser2 sont ciblées par des kinases et des phosphatases spécifiques. Cdk7, ou Kin28 chez la levure, une sous-unité du facteur de transcription II H (TFIIH), phosphoryle Ser5 à l'initiation tandis que la sous-unité Cdk9 (Ctk1 chez la levure) du P-TEFb (positive transcription elongation factor) phosphoryle les résidus Ser2 durant l'élongation (Peterlin and Price 2006). La phosphatase responsable du traitement de Ser5P est Rtr1 (Mosley, Pattenden et al. 2009) tandis que la phosphatase de Ser2P est Fcp1 (Cho, Kobor et al. 2001, Ghosh, Shuman et al. 2008). La transition d'un état Ser5P élevé à un état Ser2P coïncide avec le basculement de l'activité PolII de l'élongation lente à une élongation rapide (Phatnani and Greenleaf 2006, Ni, Saunders et al. 2008, Buratowski 2009).

Le signal poly(A) est reconnu co-transcriptionnellement par les composants du complexe de clivage et de polyadénylation. Bien que la réaction ait lieu à la fin du cycle de

transcription, certains des facteurs sont recrutés très précocement sur le complexe de transcription. Par exemple, des expériences d'immunoprécipitation ont démontrées que CPSF-73 (Ysh1p) et CstF-77 (Rna14p) sont enrichis aux sites d'initiation de la transcription (Transcription Start Sites, TSS) en plus de l'extrémité 3' (Glover-Cutter, Kim et al. 2008). Beaucoup d'autres facteurs poly(A) sont recrutés sur le CTD de la PolIII avant la maturation 3' (McCracken, Fong et al. 1997, Hirose and Manley 1998, Licatalosi, Geiger et al. 2002, Ahn, Kim et al. 2004, Zhang, Fu et al. 2005). Le CTD phosphorylé sur la sérine 2 (Ser2P-CTD) est requis pour le recrutement de beaucoup de facteurs poly(A) (Licatalosi, Geiger et al. 2002, Ahn, Kim et al. 2004). Pcf11p, en particulier, interagirait directement avec le CTD (Meinhart and Cramer 2004, Zhang, Fu et al. 2005, Zhang and Gilmour 2006). Le niveau de Ser2P diminue drastiquement jusqu'au niveau basal en aval du site poly(A), dans la même région où les facteurs poly(A) se dissocient de la machinerie de transcription (Ahn, Kim et al. 2004, Kim, Krogan et al. 2004, Ni, Schwartz et al. 2004, Zhang and Gilmour 2006, Cui, Allen et al. 2008, Garrido-Lecca and Blumenthal 2010).

En fait, la présence de la Ser2P est responsable du recrutement de la machinerie de maturation 3' (Ahn, Kim et al. 2004) dès le début de la transcription, mais une phosphorylation sur la Tyrosine 1 (Tyr1P) prévient le recrutement des facteurs de terminaison, au premier lieu desquels se trouve Pcf11p (Mayer, Heidemann et al. 2012). La sous-unité Glc7 du CPF est responsable de la déphosphorylation de Tyr1P lorsque la machinerie arrive au site de polyadénylation, permettant alors le recrutement du CF IA et la terminaison de la transcription (Schreieck, Easter et al. 2014).

La reconnaissance du signal poly(A) n'est pas essentielle uniquement pour la polyadénylation, c'est aussi un facteur clé de la terminaison de la transcription Pol II (Rosonina, Kaneko et al. 2006, Richard and Manley 2009). Il a été démontré que l'occupation par la Pol II diminue rapidement après le site poly(A) (Whitelaw and Proudfoot 1986, Logan, Falck-Pedersen et al. 1987, Connelly and Manley 1988). Il a notamment été fait preuve que la réaction de clivage, indépendamment de l'addition de la queue poly(A), est requise pour la terminaison de la transcription (Birse, Minvielle-Sebastia et al. 1998). De plus, la force d'un signal poly(A) corrèle directement avec l'efficacité de terminaison (Osheim, Proudfoot et al. 1999). Malgré des recherches intensives sur cette question, le mécanisme de terminaison de l'ARN polymérase II n'est toujours pas parfaitement caractérisé (Kim, Krogan et al. 2004, West,

Gromak et al. 2004, West and Proudfoot 2008, Ghazal, Gagnon et al. 2009, Rondon, Mischo et al. 2009). L'observation cohérente que le signal poly(A) est requis pour une terminaison efficace est interprétée comme une potentialisation de Pol II pour la terminaison lorsqu'elle a transcrit le signal poly(A) (Connelly and Manley 1988, West and Proudfoot 2008). Le changement le plus clair sur le complexe d'élongation Pol II lors de la transcription du signal poly(A) est la chute de Ser2P sur le CTD à un niveau basal. Le facteur poly(A) Pcf11p semble être un acteur clé dans le lien entre la transcription du signal poly(A) et la terminaison : Pcf11 lie le CTD et, au moins *in vitro*, entraîne la dissociation de la PolII et du transcrit naissant de l'ADN (Zhang, Fu et al. 2005, Zhang, Kim et al. 2006). De plus, l'interaction entre l'ARN et la PolII dans les étapes précoces d'association du complexe poly(A) semble amorcer l'interaction stable entre les facteurs poly(A) et la Pol II (Rigo, Kazerouninia et al. 2005).

De récentes études se sont également penchées sur un rôle de CF IA dans la réinitiation de la transcription (Ansari and Hampsey 2005, Al Husini, Kudla et al. 2013).

1.7.2. La coiffe 5'

Peu de choses sont connues, mais il a été démontré que la coiffe avait un effet positif sur la maturation 3'. La présence du Cap Binding Complex (CBC) augmente le clivage en 3', mais pas la polyadénylation. Le CBC semble également avoir un effet stabilisateur sur la machinerie poly(A) et son substrat pre-mRNA (Flaherty, Fortes et al. 1997).

1.7.3. L'épissage

L'idée que l'épissage et la maturation 3' soient couplés est apparue lorsqu'il a été découvert que des mutations sur des éléments cis de l'épissage affectaient également la polyadénylation (Cooke and Alwine 1996, Cooke, Hans et al. 1999, Vagner, Vagner et al. 2000). Il a été proposé que l'épissage proche de l'extrémité 3' d'un transcrit n'est possible qu'une fois que l'ARN est libéré de la polymérase par le clivage (Bird, Fong et al. 2005). En effet, la machinerie d'épissage et la machinerie de maturation 3' déterminent ensemble l'exon terminal (Rigo and Martinson 2008, Rigo and Martinson 2009). Ce couplage augmente clairement les probabilités d'un signal poly(A) exonique à l'extrémité 3' soit processé (Tsai, Wu et al. 1998). Les mécanismes par lesquels ce couplage aurait lieu ne sont pas bien compris, mais un nombre important d'études suggère l'implication du snRNP U1. Cet élément stimule

la maturation 3' en interagissant avec CPSF-160 (Lutz, Murthy et al. 1996). Une étude récente met également en évidence une interaction entre U2 snRNP et CPSF, et propose que cette interaction permette de coupler les deux maturations (Kyburz, Friedlein et al. 2006).

Pour les gènes sans introns, la composition du complexe d'élongation de la transcription semble jouer un rôle crucial dans la prévention d'une maturation 3' précoce.

1.8. Maladies

Nous avons vu que la maturation 3' est un processus complexe, impliquant de nombreux partenaires et est coordonnée avec de nombreuses autres activités. Elle constitue une étape clé de l'expression génique et une altération peut influencer les caractéristiques finales de l'ARNm, conduisant à une variété de maladies chez l'humain (Danckwardt, Hentze et al. 2008, Michalova, Vojtesek et al. 2013).

Le développement de tumeurs est un processus en plusieurs étapes qui résulte de l'accumulation de dysfonctionnement, les principaux étant l'activation d'un oncogène et/ou la perte d'un suppresseur de tumeur. Plusieurs études mettent en avant le rôle de mutations dans le 3'UTR conduisant à l'altération de sa longueur ou encore à la modification de séquences ciblées par des miRNAs. Certaines modifications des signaux poly(A) sur les 3'UTR de gènes spécifiques sont ainsi retrouvées associés avec différents types de cancers. La PAP a également été décrite comme étant plus active dans les cellules tumorales, le niveau d'activité de l'enzyme corrélant même avec l'agressivité de la maladie.

Le cancer est la plus connue mais pas la seule maladie à être liée à des altérations de la maturation 3'. Comme nous en avons parlé en introduction, l'étude des gènes de globine a joué un grand rôle dans la compréhension des signaux cis et l'identification des facteurs trans. Dans la α -thalassaémie, le signal AATAAA est muté en AATAAG, entraînant un défaut de maturation 3' et un 3'UTR bien plus long que la forme normale. Ce transcrit pathologique est peu compétent pour la traduction, entraînant une faible production de globine α , indispensable pour l'assemblage de l'hémoglobine (Higgs, Goodbourn et al. 1983, Orkin, Cheng et al. 1985, Whitelaw and Proudfoot 1986).

Les facteurs trans peuvent également être impliqués. Ainsi, alors qu'une maturation efficace nécessite une interaction entre CPSF-30 et la PABPN1, la protéine NS1 du virus

Influenza A perturbe cette interaction en séquestrant avec CPSF-30, ce qui conduit à une inhibition de la maturation 3' des ARNm de l'hôte, qui sont alors retenus dans le noyau. Les ARN viraux sont par contre exportés et bénéficient dans le cytoplasme d'une machine de traduction complètement disponible (Simonsen and Levinson 1983, Nemeroff, Barabino et al. 1998, Chen, Li et al. 1999, Das, Ma et al. 2008).

2. Synthèse et objectifs

La maturation 3' des ARN messagers est une étape majeure de l'expression génique, contrôlant la nature de l'extrémité 3' non-traduite synthétisée tout en participant à la régulation et la synchronisation de la valse des étapes de la synthèse de l'ARN messager. La complexité de l'interconnexion et des régulations croisées de ces mécanismes est illustrée par les hauts poids moléculaires des complexes impliqués. Ces complexes sont cependant longtemps restés méconnus, et leur composants moléculaires identifiés relativement récemment. Cependant, si certains facteurs ou domaines ont pu être résolus, aucun modèle n'est à ce jour disponible pour les complexes macromoléculaires qui sont les véritables acteurs de ces maturations.

Mon projet s'inscrit dans cette problématique. En essayant d'apporter de nouvelles connaissances, moléculaires d'une part et biologiques d'autre part, sur les composants de la machinerie de maturation de l'extrémité 3' des ARN pré-messagers. C'est dans cette optique que mon travail se concentre sur le CF IA (Clivage Factor IA), un complexe bien identifié chez la levure, indispensable aux réactions de clivage et de polyadénylation. Cependant, si les progrès effectués dans les dernières années ont apportés des éléments sur certaines structures et fonctions de domaines ou de protéines du complexe, l'organisation des sous-unités entre elles reste méconnue. Mon objectif et ambition avec ce travail est de pouvoir proposer un modèle du complexe.

Un tel modèle serait en effet un outil idéal pour étudier les régulations au niveau moléculaire et mieux comprendre les interactions complexes entre la maturation 3' et les autres processus nucléaires de l'expression génique.

En effet, il me semble que deux questions principales se posent aujourd'hui, auxquelles je tente d'apporter des éléments de réponse par ce travail.

- Comment les facteurs de l'expression génique en général, et de la maturation 3' en particulier, coordonnent leurs actions ?
- Comment les différentes sous-unités interagissent-elles ensemble, et comment ces interactions permettent une sélection précise du site de polyadénylation ?

3. Résultats

3.1. Etude des interactions du CF I

3.1.1. La reconstitution du CF IA en modèle recombinant et son application pour la cartographie des interactions

3.1.1.1. Reconstitution du CF IA et validation du modèle recombinant

Les quantités de matériel nécessaires pour les essais de cristallogénèse sont trop importantes pour que la purification du CF IA depuis des cultures de levure par la technique de l'étiquette TAP (Puig, Caspary et al. 2001) soit une option viable en laboratoire. Il est donc nécessaire de recourir à un système d'expression recombinant, afin d'exprimer les protéines d'intérêt. Cette étape peut cependant être fortement limitante si la protéine est toxique par exemple, ou si des modifications post-traductionnelles spécifiques ne sont pas disponibles chez l'organisme hôte.

Dans le cas de CF IA, la sur-expression chez la bactérie est un protocole relativement commun, utilisé dans plusieurs laboratoires depuis que les protéines ont été identifiées (Gross and Moore 2001, Bai, Auperin et al. 2007, Legrand, Pinaud et al. 2007, Noble, Beuth et al. 2007, Qu, Perez-Canadillas et al. 2007, Gordon, Shikov et al. 2011, Moreno-Morcillo, Minvielle-Sebastia et al. 2011, Paulson and Tong 2012), permettant la résolution de nombreuses structures ou des études biophysiques sur le complexe entier. Cependant, d'après la littérature, il n'existe pour le moment pas d'équipe qui travaille avec le complexe co-exprimé et co-purifié. Dans les principales études s'intéressant au complexe entier (Gross and Moore 2001, Gordon, Shikov et al. 2011), les protéines sont exprimées respectivement dans un système de cellules d'insecte et un système levure. Mais Gross et al. purifie chaque protéine séparément et reconstituent ensuite le complexe *in vitro*. De la même manière, Gordon et al. purifie Rna14p-Rna15p en complexe dans un premier temps, Pcf11p-Clp1p en complexe dans un second temps, puis reconstituent le complexe *in vitro* également.

Au laboratoire, nous nous sommes donc appliqués à développer une méthode de co-expression et co-purification du facteur CF IA entier, tout en vérifiant qu'il conservait les caractéristiques biologiques du facteur natif.

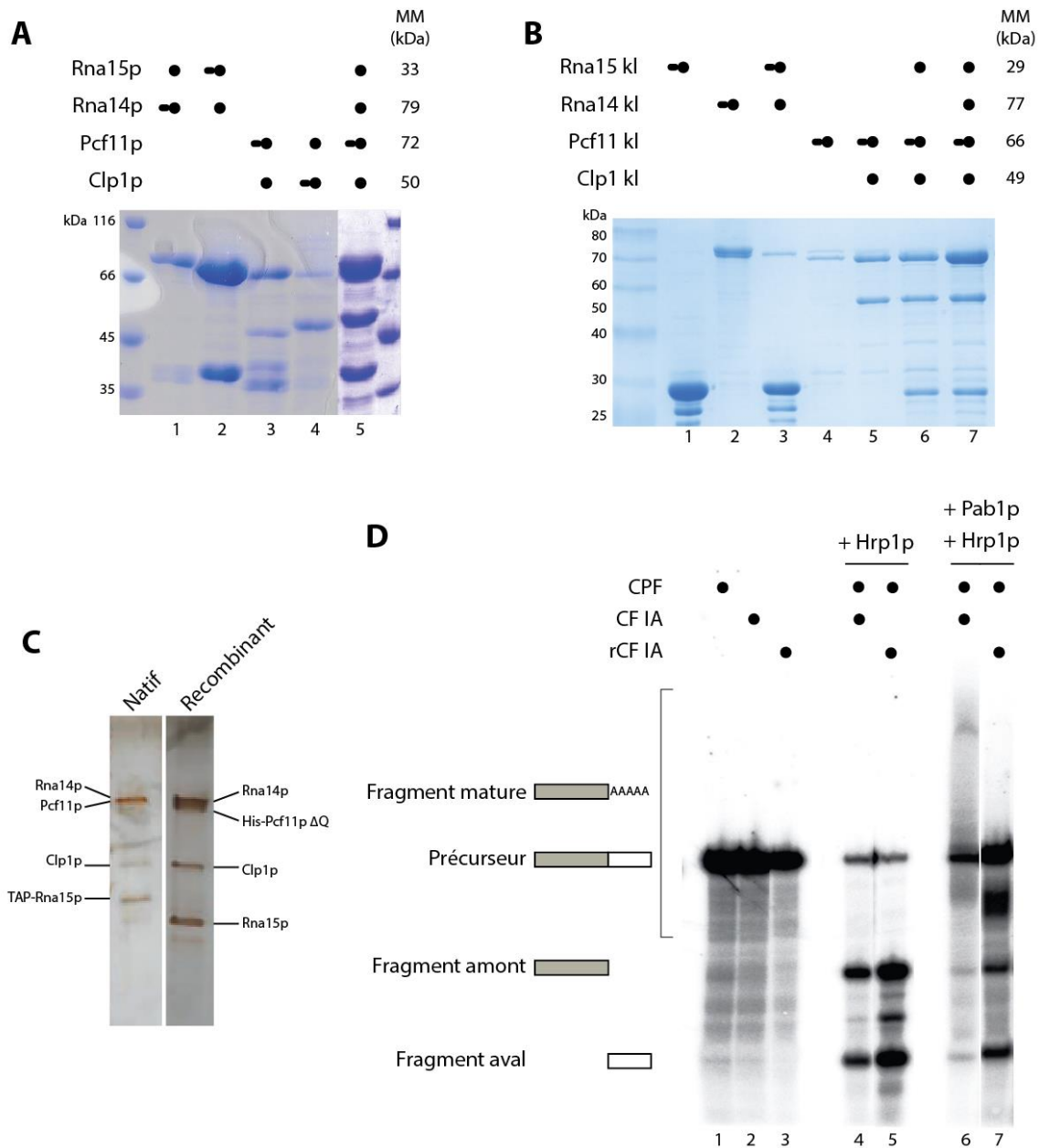


Figure 10 : Reconstitution du CF IA. Dans les panneaux A et B, MM signifie Masse Moléculaire et la forme ovale indique la sous-unité portant l'étiquette poly-His (environ 3kDa). **A.** Reconstitution du facteur CF IA de *Saccharomyces cerevisiae*. **B.** Reconstitution du facteur CF IA de *Kluyveromyces lactis*. **C.** Coloration à l'argent de facteur CF IA natif et recombinant. Réalisé par J. Guegueniat. **D.** Test d'activité des facteurs CF IA natif et recombinant. Tous les facteurs nécessaires à la réalisation du test ont été purifiés par étiquette TAP chez la levure, le CF IA recombinant est noté rCF IA. 50ng de Hrp1p sont ajoutés dans les essais 4, 5, 6 et 7, et 75ng de Pab1p sont ajoutés dans les essais 6 et 7. Essais réalisés par J. Guegueniat.

Dans le panneau A de la Figure 10, nous pouvons voir une co-expression de Rna14p et Rna15p (colonnes 1 et 2) et une co-expression de Pcf11p et Clp1p colonnes 3 et 4). Dans les deux cas, l'étiquette est alternativement portée par l'une ou l'autre des sous-unités, symbolisée par un ovale dans le diagramme au-dessus du gel et détectable grâce au léger

changement de taille. La colonne 5 présente le test d'expression réalisé sur des bactéries transformées avec tous les composants du facteur CF IA. Nous pouvons voir que nous sommes capables de distinguer toutes les sous-unités, en prenant en compte le fait que Pcf11p et Rna14p sont quasiment superposées. Le niveau d'expression est bon, relativement homogène entre les composants.

Dans la Figure 10, le panneau B représente les résultats en appliquant la même démarche au facteur CF IA de *Kluyveromyces lactis*. Les colonnes 1, 2 et 3 nous permettent de voir que l'expression de Rna14kl et Rna15kl seules ou en complexe est possible. Il en est de même pour Pcf11kl et Clp1kl, présentés dans les colonnes 4 et 5. La colonne 6 présente une expression uniquement à partir de Rna15kl, Cpl1kl et Pcf11kl, et les trois protéines se distinguent très clairement sur la migration. Dans la dernière colonne, l'expression du complexe CF IA entier est présentée. Trois bandes sont visibles, mais encore une fois, sachant la faible différence de taille entre Rna14kl et Pcf11kl en fusion avec l'étiquette, et en comparant avec la colonne 6, nous pouvons supposer que Pcf11kl et Rna14kl sont superposés.

Pour s'assurer que nous étions bien en présence du complexe entier, nous avons effectué une identification par spectrométrie de masse sur les protéines présentes dans la bande migrant aux environs de 70kDa. Les résultats montrent bien la présence de Rna14 et Pcf11, avec respectivement 82% et 57% de couverture de séquence (données non montrées).

Nous devons maintenant nous demander si le CF IA recombinant est un modèle de travail adapté, notamment en confirmant qu'il est biologiquement équivalent au facteur natif. Le panneau C montre la migration d'un facteur CF IA natif purifié par la stratégie TAP et un facteur recombinant, exprimé et purifié selon le protocole détaillé dans le chapitre 3.2.2. A l'exception des sous-unités portant les étiquettes, nous pouvons observer que les deux facteurs migrent de manière similaire en SDS-PAGE.

Pour vérifier la capacité du facteur recombinant à porter la même activité biologique que le facteur natif, nous avons effectué des essais *in vitro* avec les deux facteurs, en testant soit l'activité de clivage, soit les activités de clivage et de polyadénylation (Figure 10, panneau D). Dans ces essais, le facteur recombinant est du CF IA His Pcf11p Δ Q. Les colonnes 1, 2 et 3 sont les témoins négatifs, dans lesquels le précurseur radioactif est mis en présence des facteurs isolés. Comme prévu, aucune activité n'est détectée, et le précurseur reste intact.

Dans les essais 4 et 5, la réaction de clivage est reconstituée en mettant le précurseur en présence du CPF et du CF I (CF IA + Hrp1p). Les deux essais ont exactement le même profil. L'intensité du signal de précurseur diminue fortement, et deux bandes de masse moléculaire plus faible apparaissent. Il s'agit des deux produits de la réaction de clivage, le fragment aval et le fragment amont. Le facteur recombinant est donc parfaitement compétent dans le test de clivage *in vitro*.

Mais le CF IA est également indispensable à l'activité de polyadénylation. Pour tester cela, nous avons ajouté Pab1p à la réaction de clivage. Les résultats sont présentés dans les colonnes 6 et 7. En comparant les colonnes 4 et 6, on voit que les produits de clivage sont toujours présents et une trainée est apparue. Ces deux observations révèlent la réaction de polyadénylation effectuée avec la participation du CF IA natif. Si l'on s'intéresse maintenant aux colonnes 5 et 7, nous pouvons faire les mêmes observations, traduisant de la même manière l'efficacité de la réaction de polyadénylation en présence du CF IA recombinant. Nous pouvons cependant remarquer que le profil de la trainée est légèrement différent, avec notamment une accumulation de produits polyadénylés à une taille similaire faisant apparaître une large bande entre le précurseur et le fragment amont. Ce profil est en fait le résultat classique de la réaction (Viphakone, Voisinnet-Hakil et al. 2008), Pab1p permettant de contrôler la taille de la queue polyA synthétisée.

Le complexe CF IA peut donc être produit de manière recombinante par la co-expression de tous les partenaires, le facteur que nous obtenons à l'issue de la purification est capable de remplacer le CF IA natif dans des essais *in vitro*. Le modèle sur lequel nous travaillons au laboratoire est donc validé.

3.1.1.2. Cartographie des interactions entre Pcf11p et Rna15p/Rna14p

La présence de Pcf11p au sein de CF IA a tout d'abord été mise en évidence en tant que partenaire protéique de 70kDa, avant même que le nom de Pcf11p ne lui soit attribué (Kessler, Zhao et al. 1996). La toute première étude sur Pcf11p montre ensuite, grâce à des essais d'activité de β -Galactosidase, que la même région est responsable de l'interaction avec Rna14p et avec Rna15p (Amrani, Minet et al. 1997). L'interaction avec Clp1p n'est toujours pas connue, est Pcf11p est sommairement divisé en trois parties, parmi lesquelles le domaine central, délimité par les résidus 288 et 400, est identifié comme responsable de l'interaction

avec Rna14p et Rna15p séparément. L'interaction avec un complexe Rna14p/Rna15p n'est pas testée dans cette étude.

Depuis, étant donné son rôle central dans l'architecture de CF IA, de nombreuses autres études ont été conduites pour cartographier les interactions entre Pcf11p et les autres protéines du complexe (Gross and Moore 2001, Sadowski, Dichtl et al. 2003, Qu, Perez-Canadillas et al. 2007, Xiang, Tong et al. 2014). L'interaction de Pcf11p avec Rna14p/Rna15p en complexe a été confirmée, mais aucune n'a jamais cherché à déterminer avec plus de précision le domaine responsable de cette interaction.

Pcf11p est également un des facteurs les moins bien caractérisé du CF IA. Il est bien connu cependant que c'est un facteur avec deux fonctions indépendantes qui peuvent être découplées *in vivo* (Sadowski, Dichtl et al. 2003). La partie N-terminale est impliquée dans la terminaison de la transcription via l'interaction avec le CTD tandis que la partie C-terminale agit dans la maturation 3' en tant que membre du CF IA. Il est donc tout à fait intéressant pour nous d'identifier la borne N-terminale de la plus petite construction de Pcf11p permettant une interaction avec Rna14p et Rna15p.

La Figure 11, panneau A, montre la capacité de différentes constructions de Pcf11p à interagir ou non avec Rna14p et Rna15p. Ainsi, tandis que la protéine complète est bien capable d'interaction, le domaine 1-233 caractérisé comme étant le CID n'en n'est pas capable. Nous pouvons voir également qu'une construction commençant au résidu 298 est toujours capable de former cette interaction, de même d'une protéine débutant au résidu 331. Finalement, nous pouvons voir qu'un Pcf11p réduit aux résidus 331 à 417 est toujours compétent pour l'interaction avec Rna14/Rna15p.

Cependant, Rna14p et Rna15p interagissant eux-mêmes entre eux, nous ne pouvons savoir si ce fragment de Pcf11p n'a pas perdu l'interaction avec l'une des deux protéines mais que cela soit compensé par son interaction avec la seconde. Pour tester cela, nous avons réalisé des tests d'interaction entre Pcf11p (331-417) et les protéines seules ou les protéines exprimées ensembles (Figure 11, panneau B), dans deux souches d'expression différentes.

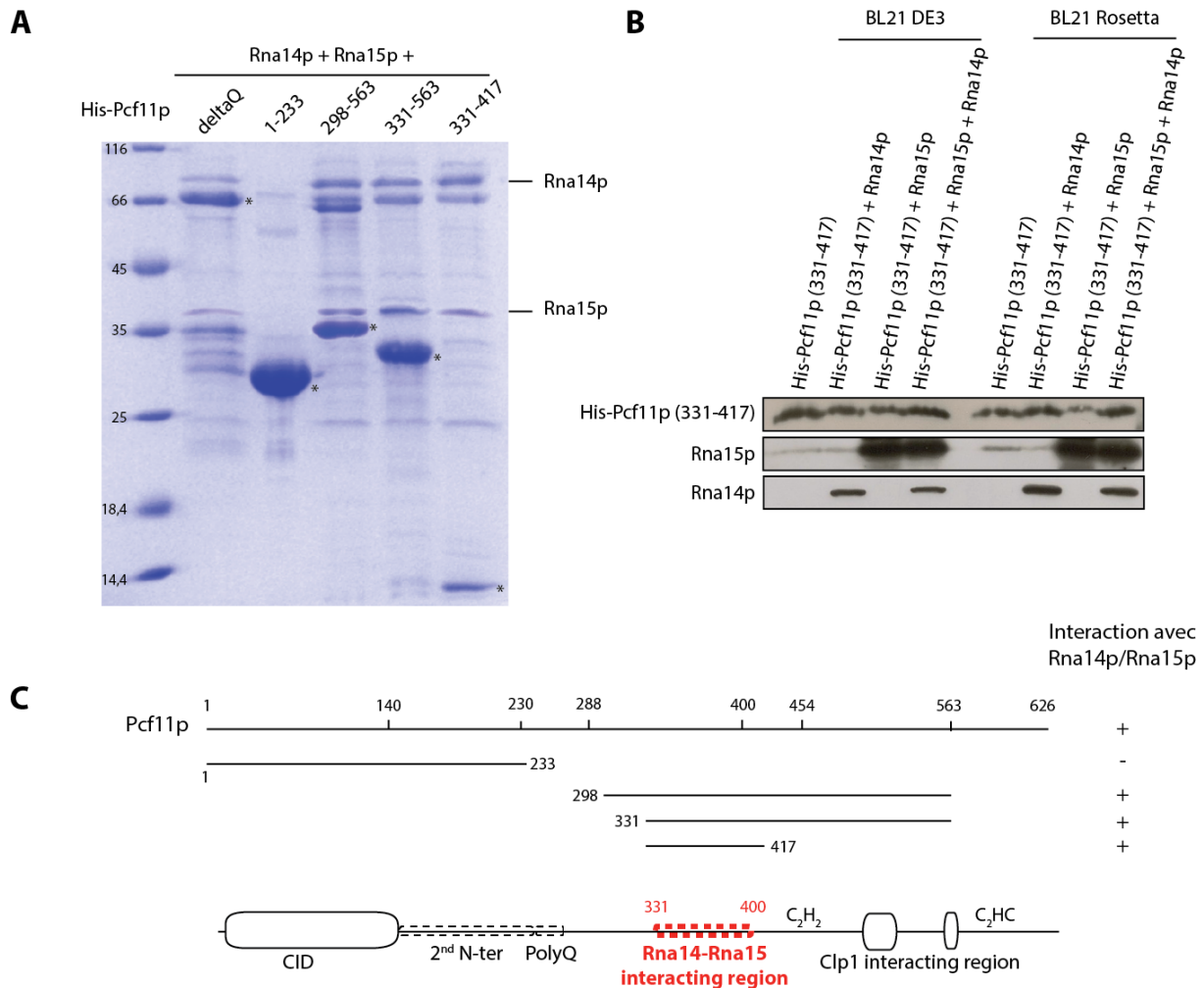


Figure 11 : Cartographie de l'interaction entre Rna14p, Rna15p et Pcf11p. **A.** Co-purification des sous-unités Rna14p et Rna15p par différentes constructions de Pcf11p. Les fragments de Pcf11p sont indiqués par un astérisque (*) sur le gel. **B.** Western blot effectué sur des essais de co-purification. Pcf11p (331-417) est révélé avec un anticorps polyclonal anti His tandis que Rna14p et Rna15p sont révélées par des anticorps polyclonaux spécifiques. **C.** Diagramme résumant les constructions de Pcf11p interagissant ou non avec Rna14p et Rna15p, permettant de préciser la région de l'interaction.

Nous observons que Pcf11p (331-417) interagit avec Rna14p et Rna15p individuellement. Cela signifie donc que le domaine 331-417 de Pcf11p est compétent pour l'interaction avec les protéines seules, comme le domaine 288-400 utilisé par Amrani et al. (Amrani, Minet et al. 1997). Nous pouvons donc affiner le domaine minimum d'interaction de Pcf11p avec les protéines Rna14p et Rna15p pour le réduire à la région 331-400 (Figure 11, panneau C).

Malgré cet affinement des régions d'interactions, et à cause de l'absence totale d'étude spécifique à ce sujet, l'interaction entre Pcf11p, Rna14p et Rna15p reste très méconnue. Les résidus impliqués ne sont pas connus, et aucune des études de mutants chez la levure menées à ce jour n'a mis en évidence de mutants pour cette interaction. Par contre,

de manière assez curieuse, des mutants de Clp1p perturbent l'interaction entre Pcf11p et Rna14p/Rna15p (Haddad, Maurice et al. 2011, Dupin and Fribourg 2014). Cette question a également été explorée au laboratoire et sera discutée dans le chapitre 3.5.

3.1.1.3. Essais sur la dimérisation de Rna15p

Il est parfaitement établi que Rna14 peut dimériser (Bai, Auperin et al. 2007, Legrand, Pinaud et al. 2007, Paulson and Tong 2012). L'interaction entre Rna14p et Rna15p est également bien documentée (Moreno-Morcillo, Minvielle-Sebastia et al. 2011, Paulson and Tong 2012), ce qui a très rapidement conduit à considérer la présence de deux sous-unités de Rna15 au sein du CF IA, comme conséquence de la dimérisation de Rna14. Aucune confirmation directe n'a cependant encore été reportée. Dans un même temps, les homologues mammifères de Rna14p et Rna15p, CstF-77 et CstF-64, font partie d'un complexe dont le dernier partenaire, CstF-50, est capable de dimériser également (Moreno-Morcillo, Minvielle-Sebastia et al. 2011). Ces différents points, ainsi que des résultats obtenus durant ma thèse (cf chapitres 3.2.1.2.1 et 3.2.1.3) m'ont amené durant ma thèse à vouloir interroger la stœchiométrie de Rna15p au sein du complexe CF IA de manière directe (Chapitre 3.2.1, page 94). J'ai également voulu lever les doutes sur une possible homodimérisation de Rna15p.

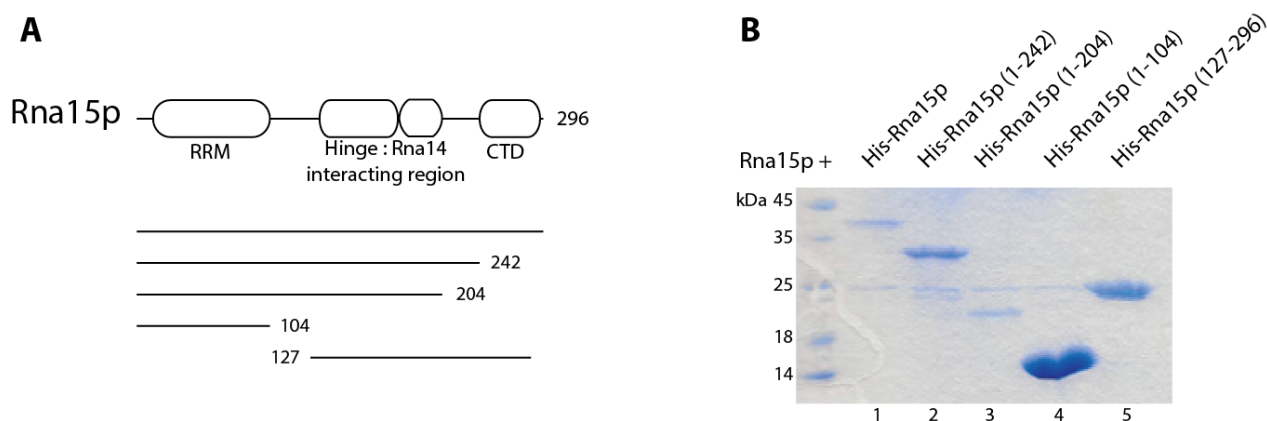


Figure 12 : Test de dimérisation de Rna15p. **A.** Diagramme des constructions utilisées pour essayer de co-purifier un Rna15p entier. **B.** Essais de co-purification de Rna15p entier par de domaines de Rna15p en fusion avec l'étiquette His.

Le panneau A présenté dans la Figure 12 illustre les différentes constructions testées dans cette expérience de co-purification. Le panneau B présente les résultats obtenus. Si l'on voit clairement la construction « appât » dans chaque essai, aucune bande n'est jamais visible au niveau du marqueur de 35kDa, zone de migration de Rna15p non étiquetée. Nous pouvons

donc sérieusement penser que Rna15p n'est pas capable d'homodimériser en l'absence d'autres sous-unités du complexe. L'objection pourrait être faite qu'il est possible qu'une homodimérisation soit présente entre les fragments portant l'étiquette, biais qui serait lié à la proximité des protéines lors de leur traduction, et qui serait alors non détectable par cette technique. Cela reste toujours possible, cependant, de nombreux exemples montrent que ce biais n'est pas suffisamment important pour empêcher l'homodimérisation d'être détectée dans le cas d'autres protéines, comme Rna14p ou CstF50 au laboratoire. Il est donc raisonnable de penser qu'il ne sera pas non plus à même de nous empêcher de détecter une homodimérisation de Rna15 si elle devait exister.

3.1.2. Caractérisation des interactions *in vitro*

Utilisés conjointement en routine dans les essais fonctionnels ou beaucoup étudiés séparément, seules deux études récentes se concentrent sur la caractérisation biophysique des interactions du CF I *in vitro*. Cependant, Leeper et al. (Leeper, Qu et al. 2010), qui explorent la liaison à l'ARN, n'utilisent que les RRM de Hrp1p et un RRM de Rna15p, se situant donc loin du contexte du CF IA. Barwal et al. (Barwal, Lee et al. 2012) vont un peu plus loin en intégrant Rna14p dans leurs essais, et remarquent déjà des différences. Il semble donc tout à fait pertinent d'explorer les interactions avec un CF IA entier.

3.1.2.1. Résonance plasmonique de surface

Une technique reconnue pour les études d'interaction entre deux partenaires est la résonance plasmonique de surface ou SPR. Dans cette technique, l'un des partenaires est mis en présence de l'autre partenaire précédemment fixé sur une surface d'or. S'il y a une interaction, la masse présente sur la surface va donc être modifiée, ce qui pourra être suivi en mesurant la variation de l'indice de réfraction de la lumière par la surface. La SPR détecte donc une variation de masse.

La principale limite de la SPR réside dans la nécessité d'avoir un des partenaires fixé au support. Pour des protéines, la fixation peut se faire de manière aspécifique en activant la surface (formation d'extrémités réactives COOH) qui sera alors capable de former des liaisons covalentes. Pour fixer de l'ARN, il est en revanche nécessaire d'utiliser des ARN couplés à la biotine et de fixer de la streptavidine sur la surface. Comme nous nous intéressons au CF I (CF

IA + Hrp1), nous avons contrôlé que nos trois partenaires, l'ARN, le complexe CF IA et les RRM de Hrp1p n'avaient pas d'interaction aspécifique avec la surface. L'ARN et Hrp1p se sont très bien comportés, mais malheureusement, malgré le test de différentes puces (SAHC®, CM5® ou SA®) le CF IA collait à l'interface, créant un bruit de fond très important. Ce comportement fut inattendu, car des essais effectués avec Rna14p-Rna15p indiquaient que ces protéines ne génèrent pas de bruit de fond (Noble, Walker et al. 2004). Cela signifie que dans tous nos essais impliquant le complexe, c'est lui qui devra être immobilisé. Et c'est une mauvaise nouvelle car comme nous l'avons dit, la SPR détecte en fait une variation de masse. Il est donc beaucoup plus facile de détecter la fixation de 300 kDa sur un 42 mer (environ 13 kDa, soit une augmentation de 2300%) que la fixation du 42mère sur 300 kDa (soit une augmentation de 4%). De plus, devoir immobiliser le complexe n'est pas intéressant en termes de stratégie, car le protocole de SPR comprend une étape de régénération du « ligand », afin de pouvoir faire plusieurs expériences avec le même matériel, permettant de l'économiser mais surtout de comparer plus facilement les résultats. Si les protéines seules résistent en majorité bien à ce traitement, il y a de forte chance que le complexe se dissocie.

Le panneau A de la Figure 13 représente la construction de Hrp1p utilisée dans cette expérience dans le contexte de la protéine entière. La séquence du site pA de Cyc1sc est également représentée en regard du site pA consensus de la levure.

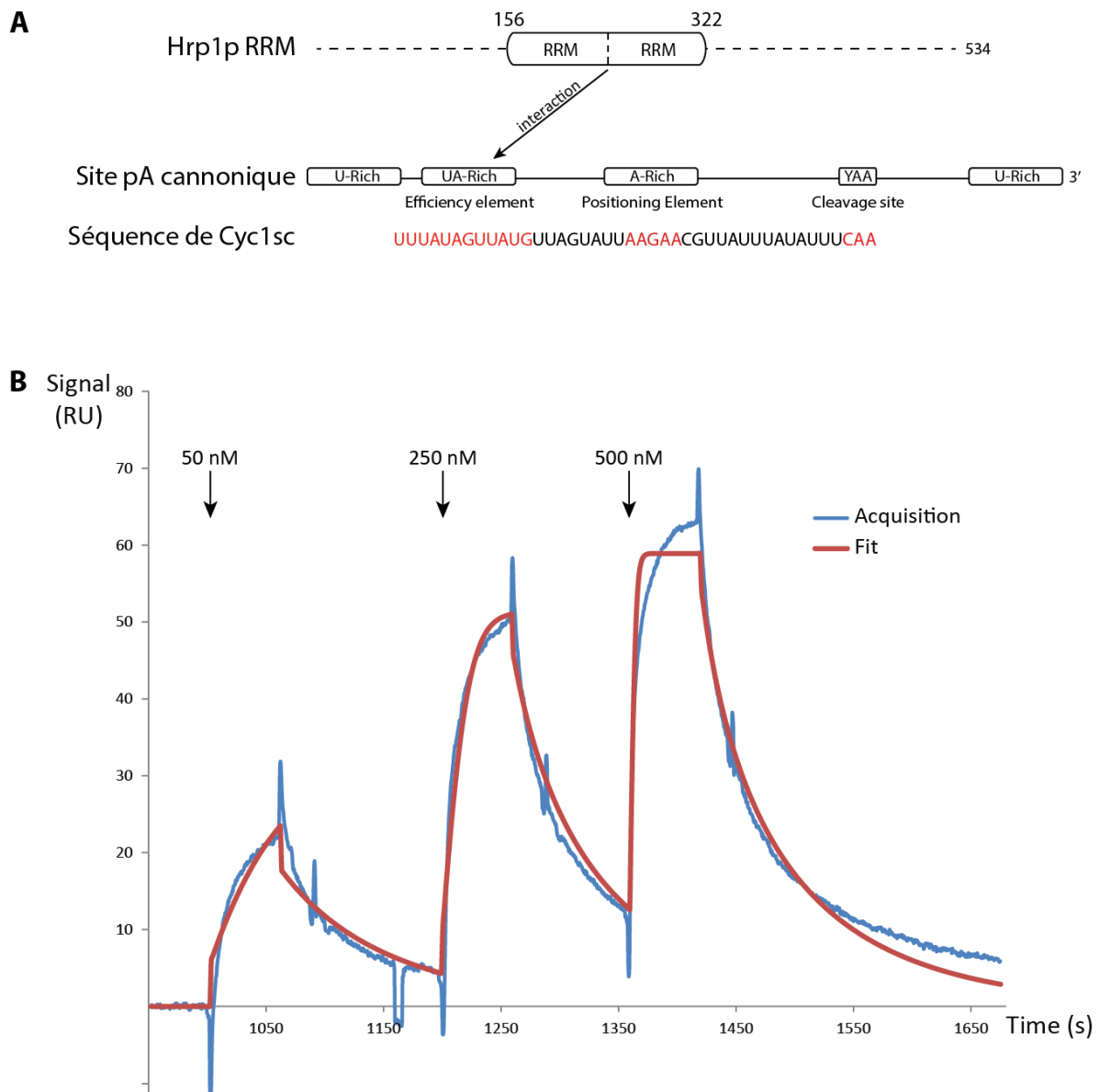


Figure 13 : Analyse cinétique de la liaison de Hrp1p à l'ARN. **A.** Schéma des constructions de Hrp1 et de l'ARN Cyc1sc utilisés durant l'expérience. **B.** Les essais ont été réalisés à 20°C, 20µl/min. 400 RU de Hrp1 (156-322) ont été immobilisés sur la puce et 20µl d'ARN sont injectés séquentiellement, à concentration croissante. Le fit est réalisé un modèle de Langmuir 1 :1.

Dans un premier temps, nous nous sommes intéressés à l'interaction entre les RRM de Hrp1 immobilisés et l'ARN (Figure 13, panneau B). Cette liaison est documentée dans la littérature (Valentini, Weiss et al. 1999, Perez-Canadillas 2006, Leeper, Qu et al. 2010), mais j'ai tenu à la tester néanmoins avant d'envisager le complexe CF IA puis le complexe CF I dans son ensemble. L'allure de la courbe traduit une constante d'association assez forte, ainsi qu'une tendance à la réassociation, ce qui nuit à la qualité de la courbe modèle. Ces caractéristiques déjà décrites par Perez-Canadillas et al. sont très certainement liées à la présence de deux RRM. Nous avons réussi à calculer une constante d'association pour le

complexe, que nous établissons à 27 μ M. Très faible, elle traduit une excellente association entre les deux partenaires, et est très comparable aux valeurs obtenues par Valentini et al. (2 à 20 μ M). Un ARN étranger, le miR-122, avec lequel aucun signal n'a été enregistré, a été utilisé comme contrôle de la spécificité de l'interaction (données non montrées).

Je me suis ensuite intéressé à l'interaction entre le CF IA et l'ARN. Compte tenu des limites exprimées précédemment, le CF IA a été fixé sur une puce CM5 afin de tester la réponse à l'injection d'ARN. Comme nous nous y attendions, la différence de masse générée par la fixation de l'ARN fut trop faible pour enregistrer un signal analysable, et le bruit de fond est extrêmement important, certainement dû à une dissociation/association du complexe.

Nous ne pouvons malheureusement pas outrepasser cette limite technique. Il n'a encore jamais été fait mention dans la littérature d'étude sur des complexes stabilisés de manière covalente. Et quand bien même, nous n'aurions pas su quel crédit apporter à des constantes d'interactions obtenues dans ces conditions. J'ai alors fait quelques essais exploratoires en ITC (Titration Calorimétrique Isotherme), mais sans parvenir à obtenir plus d'informations (données non montrées).

3.1.2.2. Analyse du CF I couplé à l'ARN en retard sur gel

Nos différentes approches cinétiques ne nous permettant finalement pas d'explorer autre chose que l'interaction entre Hrp1p et l'ARN, je me suis tourné vers des essais de retard sur gel. Cette approche a déjà été utilisée dans l'étude de la maturation 3' des ARNm, avec la caractérisation de l'interaction entre le CPF et l'ARN notamment (Keller, Bienroth et al. 1991), mais jamais sur le complexe CF IA et *a fortiori* sur le complexe CF I. Par contre, en 2004, des essais avec des protéines recombinantes ont été effectués afin de titrer l'interaction entre Rna15p ou Rna15p-Rna14p avec l'ARN (Noble, Walker et al. 2004).

Je me suis donc inspiré de leur approche pour faire des essais visant à titrer les interactions du complexe CF I avec l'ARN, dans l'objectif d'identifier des conditions optimales d'assemblage *in vitro* afin de reconstituer le complexe pour des essais de cristallogénèse. Les constructions sont présentés dans le panneau A de la Figure 14.

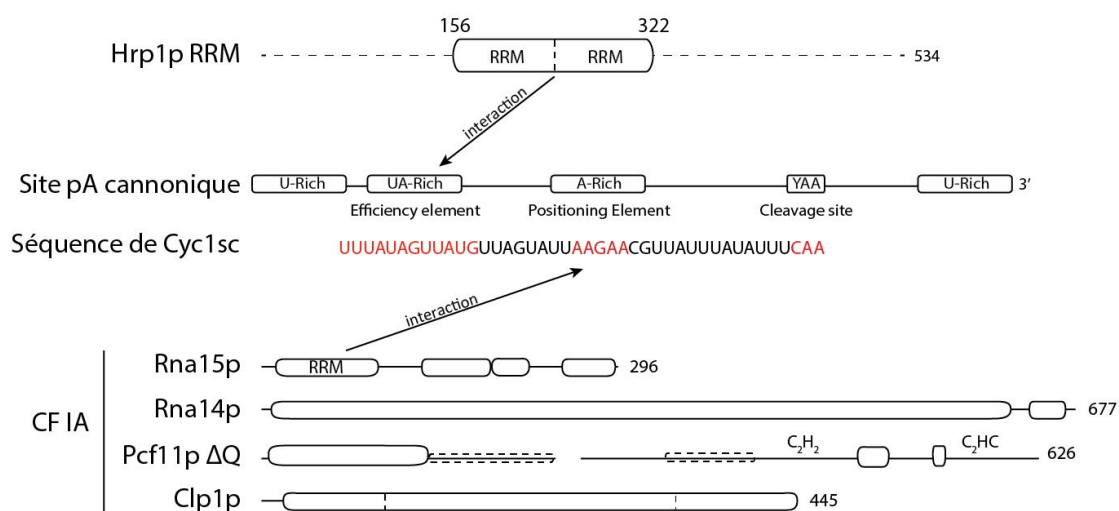
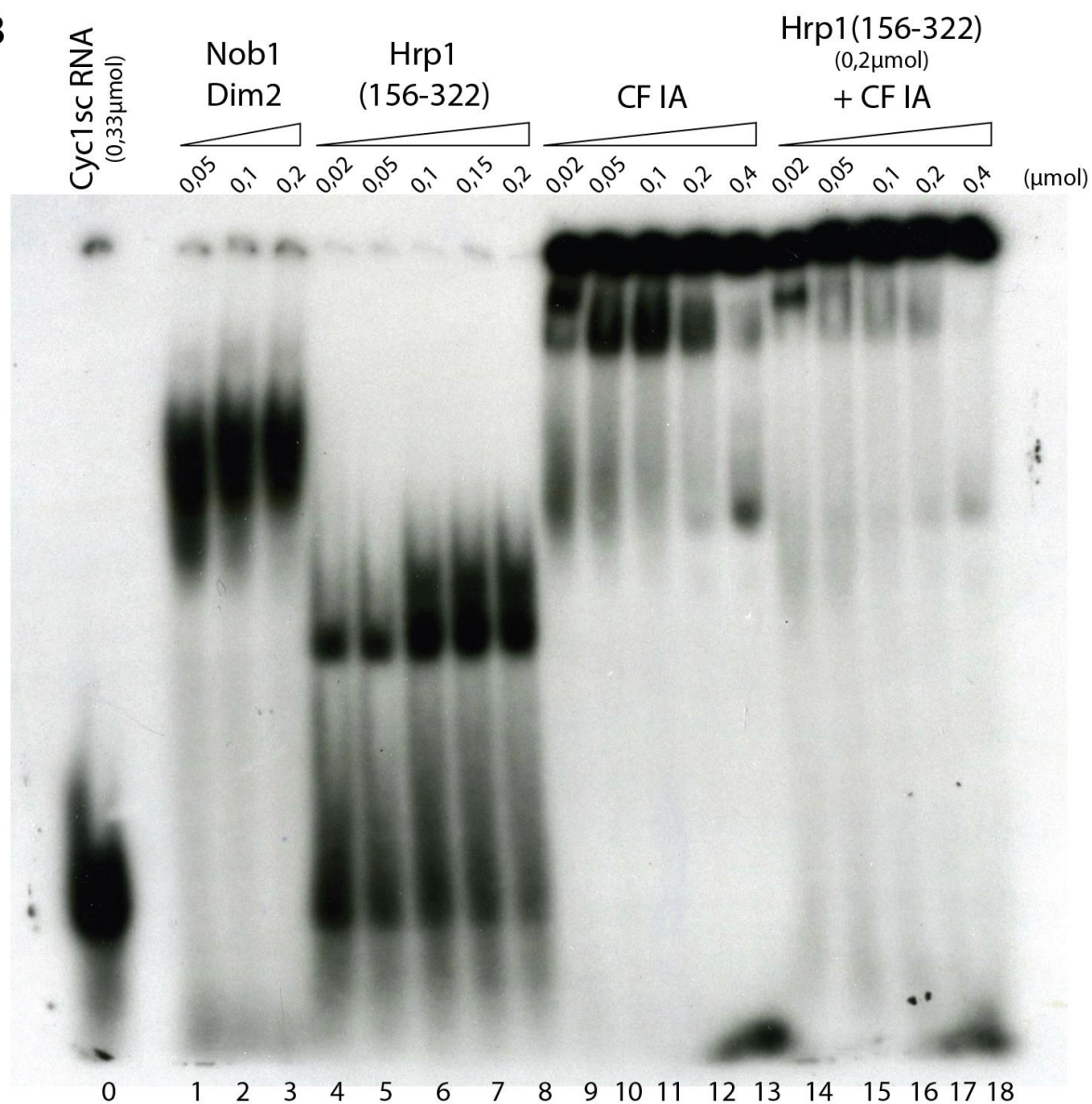
A**B**

Figure 14 : Essais de retard sur gel pour explorer les interactions du CF I. **A.** Schéma des facteurs protéiques et de l'ARN Cyc1sc utilisés durant l'expérience. **B.** Les partenaires sont incubés avec 0,33 μ mol de sonde radioactive pendant 20 minutes à 4°C puis résolues sur un gel de polyacrylamide à 4,5% en conditions natives. 0,05, 0,1 et 0,2 μ mol du complexe Nob1-Dim2 sont utilisées en contrôle. Les quantités de Hrp1 (156-322) utilisées sont respectivement de 0,02, 0,05, 0,1, 0,15 et 0,2 μ mol pour chaque essai tandis qu'elles sont de 0,02, 0,05, 0,1, 0,2 et 0,4 μ M pour le CF IA Δ Q.

Le panneau B présente la révélation de la migration des essais de formation des complexes. Le puits 0 est un contrôle avec l'ARN seul tandis que les puits 1 à 3 sont des contrôles positifs avec le complexe Nob1-Dim2 recombinant, également étudié au laboratoire et connu pour lier l'ARN (Lamanna and Karbstein 2009). Les conditions 4 à 8 sont des essais d'interaction entre Hrp1p (156-322) et l'ARN. Comme nous nous y attendions, un complexe est formé et nous voyons que le signal de l'ARN seul est fortement atténué tandis qu'une nouvelle bande apparaît. De plus, l'intensité de cette bande est concentration dépendante. A la quantité maximale, 0,2 μ mol de Hrp1p sont mis en présence de 0,33 μ mol d'ARN. Le complexe ayant une stœchiométrie 1 :1, nous nous attendons à être en excès d'ARN, ce qui est cohérent avec le fait que le signal de l'ARN seul ne disparaisse pas.

Dans les puits 14 à 18 et 9 à 13, le complexe CF IA est mis en présence de l'ARN, respectivement avec ou sans l'ajout d'Hrp1p. Dans les deux cas, on doit remarquer qu'une partie importante du signal ne sort pas du puits, et ce quel que soit le temps de migration (donnée non montrée). Il semblerait que nos conditions expérimentales entraînent la précipitation du complexe dans le puits, ou alors un réseau est formé entre l'ARN et les protéines, limitant pour l'entrée des objets dans le gel. Lorsque Hrp1p n'est pas présent, nous pouvons cependant observer l'apparition d'une bande. Sans que je puisse l'expliquer, la présence d'un second signal, correspondant à un objet migrant plus loin dans le gel, mais moins que Hrp1p-ARN, semble corrélée avec les variations d'intensité du premier signal. Lorsque Hrp1p est ajouté, nous retrouvons ces deux signaux, mais à des intensités plus faibles, laissant supposer qu'encore plus de matériel est resté dans les puits.

Si CF IA (ou CF I) lié à l'ARN peut-être une proposition d'interprétation pour le signal le plus proche des puits, nous n'avons aucun moyen d'en être convaincu. Toutes les hypothèses sont possibles pour le second signal, bien moins intense. Il est envisageable qu'il s'agisse de Rna15p liée à l'ARN, au vu de la situation par rapport aux complexes formés par Hrp1p (156-322) d'une part et Nob1-Dim2 d'autre part.

Observer la quasi-totalité du signal bloqué dans le puits fut une surprise, car Noble et al. utilisent des concentrations en Rna14p-Rna15p plus importantes sans observer d'effet similaire. Afin d'essayer d'améliorer la migration, j'ai tenté de migrer plus longtemps, sans résultats, mais également de réduire la réticulation du gel, afin de potentiellement permettre l'entrée à des objets plus important, sans résultats également (données non montrées).

3.2. Etude biochimique de CF IA

3.2.1. Stratégies & constructions

L'étude structurale d'un complexe de 300 kDa est un projet ambitieux qui doit être abordé sous plusieurs angles. Pour cette raison, deux stratégies ont été suivies. D'une part, une exploration des homologues du complexe chez d'autres espèces, qui pourraient avoir des comportements plus intéressants en expression, purification et cristallogénèse, notamment en terme de stabilité ou de dégradation. D'autre part, l'expression de sous complexes et de protéines tronquées afin de minimiser l'impact des régions non structurées et/ou mobiles.

Pour cela, j'ai débuté en faisant une analyse rapide par des services de prédiction tels que GlobPlot (Linding, Russell et al. 2003), IUPred (Dosztanyi, Csizmok et al. 2005) ou encore HHpred (Soding, Biegert et al. 2005) des protéines composant le CF IA chez *Saccharomyces cerevisiae*, *Kluyveromyces lactis* et *Schizosaccharomyces pombe*. La Figure 15 regroupe les prédictions de GlobPlot, qui sont les plus globales et ne s'intéressent qu'à la propension d'une séquence d'acides aminés à être ordonnée (traduit par la modélisation par du repliement globulaire) ou non.

Si nous comparons les protéines entre elles, nous voyons sans surprise que Rna15 et Pcf11 sont les séquences avec les plus de régions prédites désordonnées. Entre les trois espèces, nous remarquons immédiatement que les homologues de *Schizosaccharomyces pombe* sont systématiquement plus longs et prédits plus désordonnés. Pour cette raison, la plupart du travail présenté dans cette thèse sera effectué avec les protéines de *S. cerevisiae* et *K. lactis*.

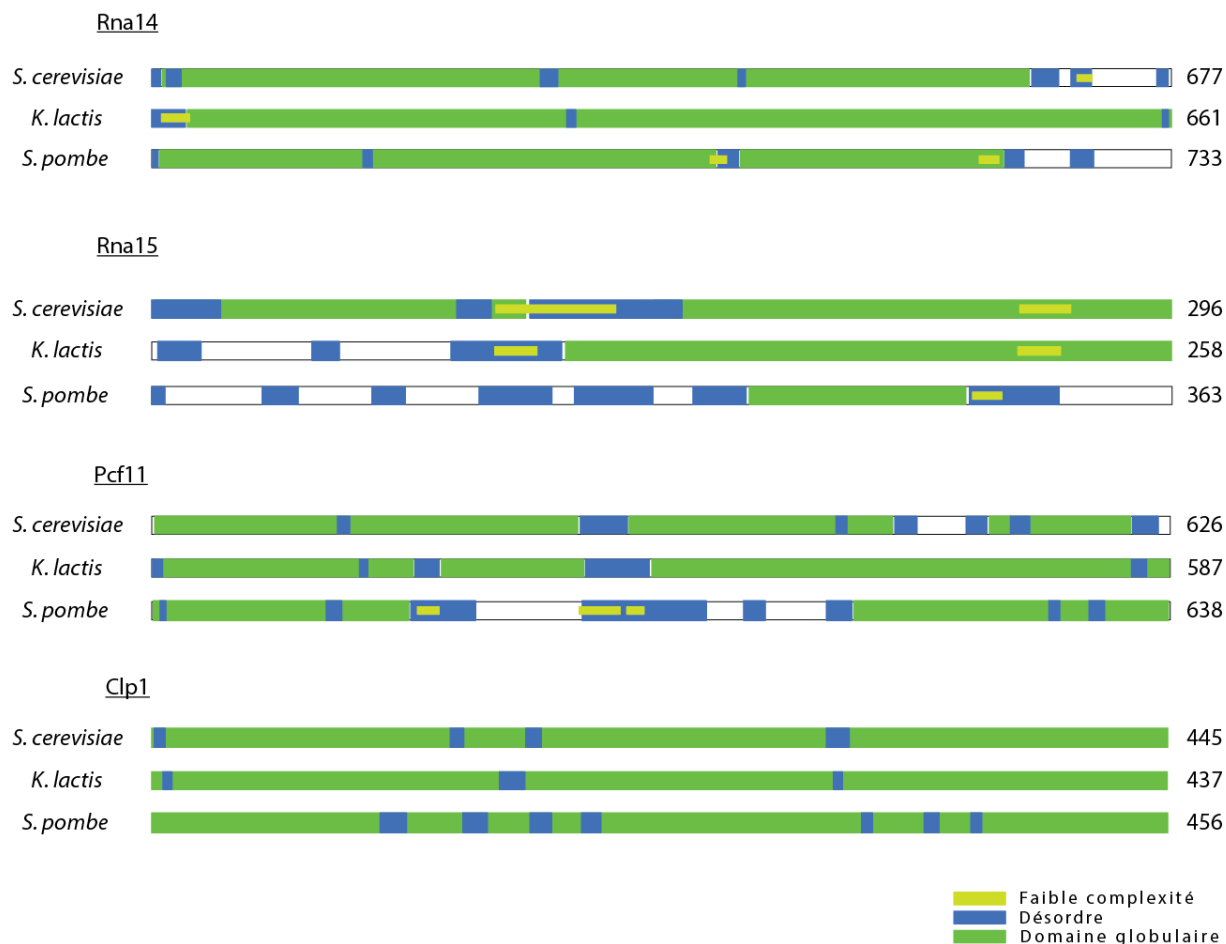


Figure 15 : Représentation schématique des prédictions de tendance à l'ordre ou au désordre le long des séquences selon GlobPlot.

Grâce à ces informations et aux informations issus de la cartographie des interactions (cf La reconstitution du CF IA en modèle recombinant et son application pour la cartographie des interactions, p.33),

nous avons sélectionné les constructions les plus intéressantes à explorer. Rna14 et Clp1 étant presque entièrement ordonnés, je n'ai pas cloné de construction tronquée de ces protéines. Par contre,

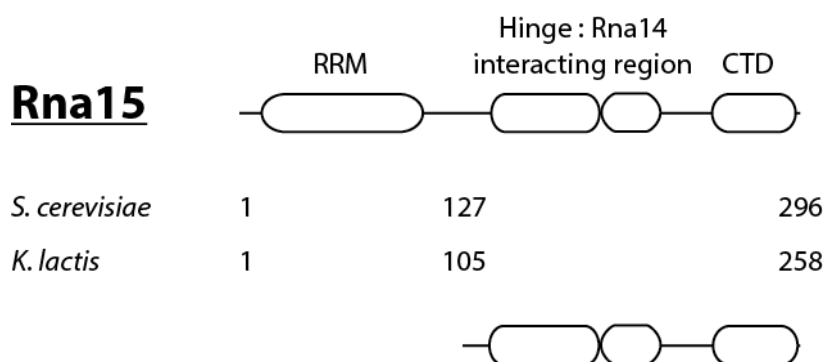


Figure 16 : Représentation des deux principales constructions de Rna15 utilisées. Les surfaces blanches représentent les domaines dont la structure est résolue. La concordance entre les séquences de *S. cerevisiae* et *K. lactis* est indiquée.

plusieurs séquences de Rna15 (Figure 16) et de Pcf11 (Figure 17) ont été utilisées durant cette thèse.

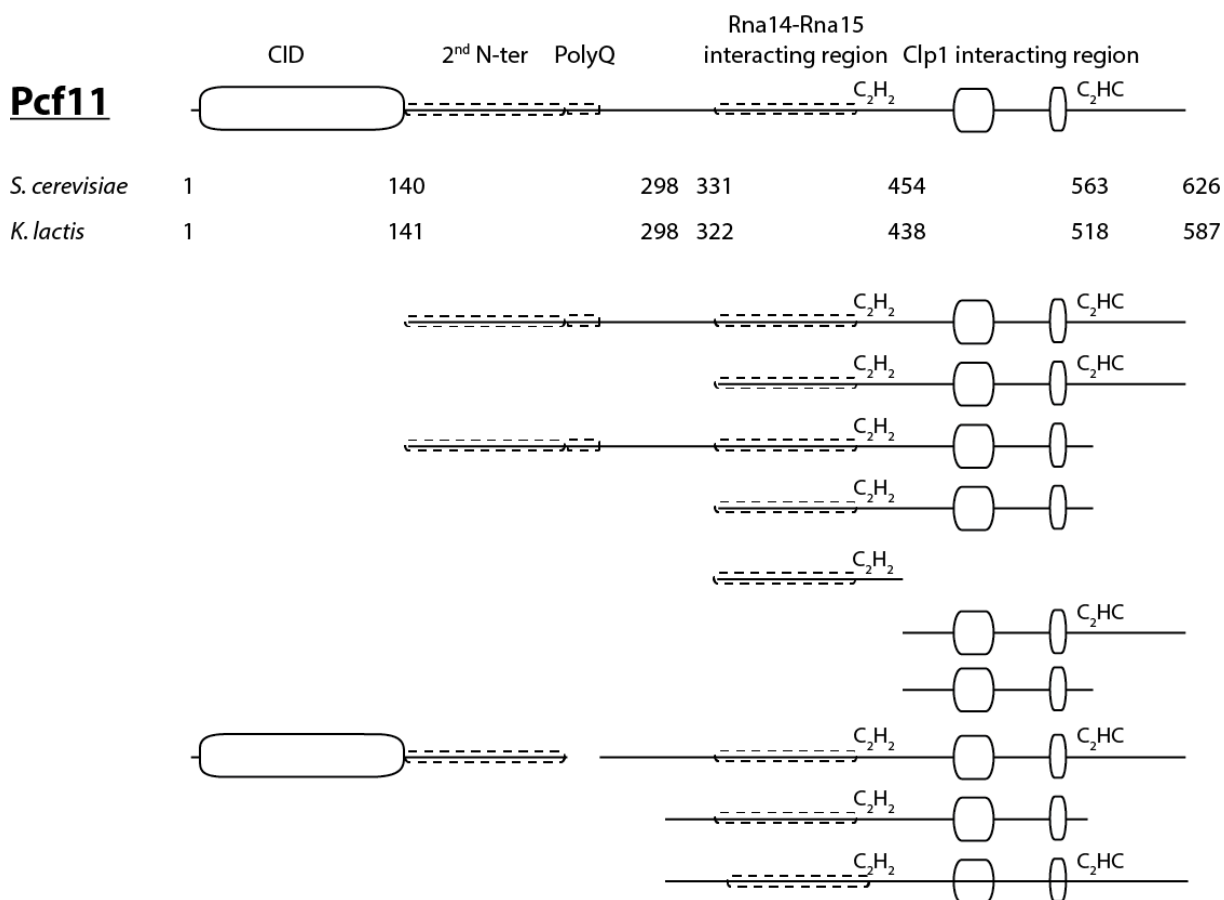


Figure 17 : Représentation des principales constructions de Pcf11 utilisées. Les surfaces blanches représentent les domaines dont la structure est résolue. La concordance entre les séquences de *S. cerevisiae* et *K. lactis* est indiquée.

Par le jeu de ces constructions, notre objectif sera d'essayer de réduire au maximum les régions non ordonnées présentes dans le complexe, et donc diminuer la mobilité.

Les références aux échantillons de CF IA seront faites en citant la sous-unité portant l'étiquette ainsi que les sous-unités troquées. Les sous-unités non mentionnées sont sous-entendues présentes et entières.

3.2.2. Expression et purification

La stratégie que nous adoptons au laboratoire pour exprimer CF IA n'est utilisée par aucune autre équipe travaillant sur le sujet. Dans toutes les études publiées sur le facteur CF IA, le complexe n'est en effet jamais produit par la coexpression de toutes les sous-unités (Gordon, Shikov et al. 2011). Les sous-unités ou les sous-complexes sont exprimés et purifiés

séparément, puis le complexe est reconstitué *in vitro*. Au laboratoire, nous utilisons la coexpression polycistronique afin d'exprimer et de purifier le complexe assemblé dans le cytoplasme bactérien.

La coexpression, et la coexpression polycistronique plus particulièrement, est une technique reconnue pour l'assemblage des complexes multiprotéiques, atténuant les potentiels problèmes de stoechiométrie et de dégradation tout en améliorant les rendements de l'expression (Diebold, Fribourg et al. 2011) et la solubilité des complexes. Durant ma thèse de doctorat, j'ai exclusivement purifié CF IA ou ses sous-complexes à partir de bactéries coexprimant les protéines, à partir de vecteurs polycistroniques dans la mesure du possible.

Les schémas de la Figure 18 à la Figure 23 présentent les stratégies de purification employées pour les expressions les plus couramment réalisées durant ma thèse. L'objectif de ces méthodes est d'obtenir un matériel homogène de pureté optimale, conformément aux impératifs de la cristallogénèse. Une purification classique de CF IA sera effectuée à partir de deux litres de culture et permettra la purification d'environ un à trois milligrammes de complexe.

Rna14 purification Flowchart

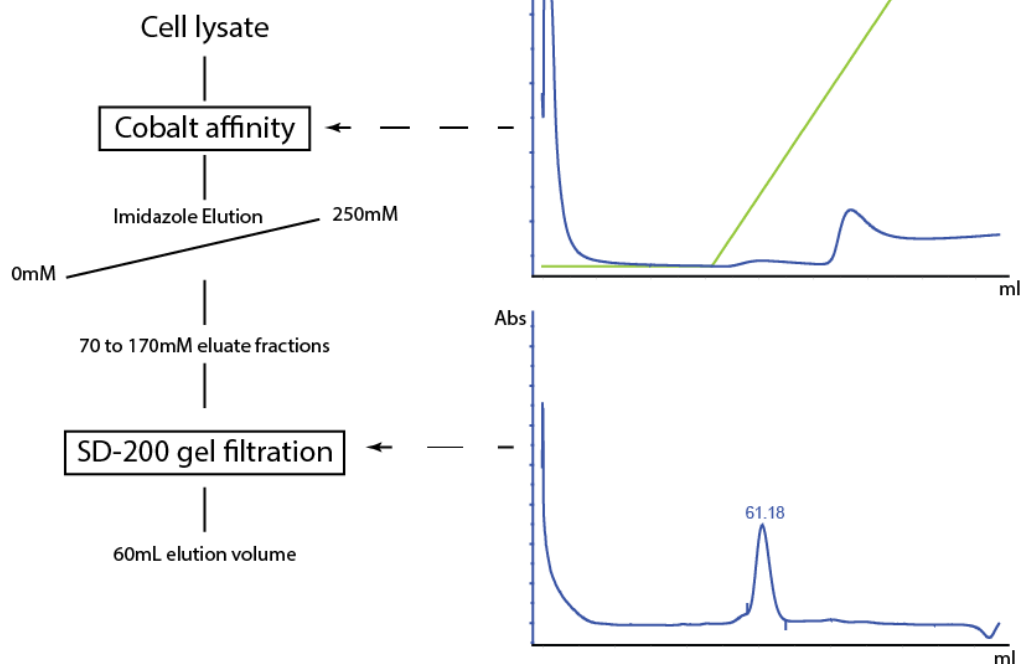


Figure 18 : Diagramme type de la purification de His-Rna14p. Pour chaque étape, la courbe en vert représente le gradient d'élution, la courbe en bleue l'absorbance à 280 nm. Le volume indiqué à l'étape de filtration sur gel est le volume d'élution du pic.

La purification de Rna14p est effectuée en deux étapes. Une première étape d'affinité contre l'étiquette est effectuée en Tris 25mM pH 7,5 et NaCl 150mM à 5 ml/min, l'élution se faisant grâce à un gradient d'imidazole. Après analyse sur un gel SDS-PAGE à 12%, les fractions d'intérêt sont concentrées et chargées pour une étape de filtration sur gel sur une résine Superdex-200 dans un tampon Tris 25mM pH 7,5, NaCl 250mM et DTT 1mM, à 0,5 ml/min. Un pic unique, correspondant à une dimérisation du monomère His-Rna14p sera élué à environ 60 ml.

Rna15 purification Flowchart

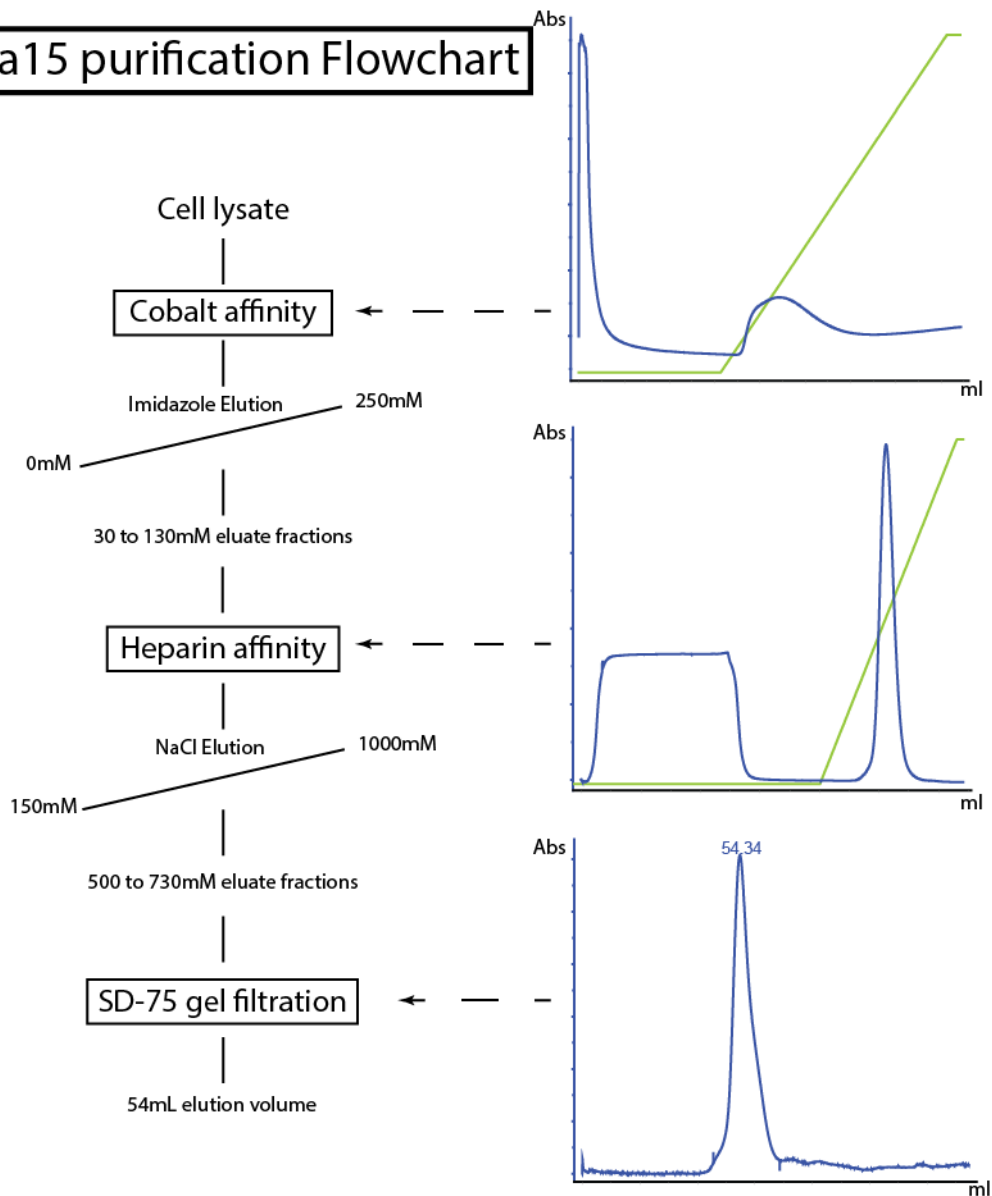


Figure 19 : Diagramme type de la purification de His-Rna15p. Pour chaque étape, la courbe en vert représente le gradient d'éluion, la courbe en bleue l'absorbance à 280 nm. Le volume indiqué à l'étape de filtration sur gel est le volume d'éluion du pic.

La protéine His-Rna15p recombinante peut être purifiée par une première étape d'affinité avec de la résine Co^{2+} . Cette affinité se fait dans un tampon Tris 25mM pH 7,5 et avec 150mM de NaCl. Les fractions sont contrôlées sur un gel SDS-PAGE 15% et chargées sur une colonne Heparine. Une éluion à pH 7,5 (Tris 25 mM) par un gradient de NaCl permet de récupérer la protéine. Les fractions sont alors concentrées et chargées sur de la résine Superdex 75 pour une exclusion de taille. Le monomère de protéine sort de la colonne à 54ml d'éluion.

Rna14/Rna15 purification Flowchart

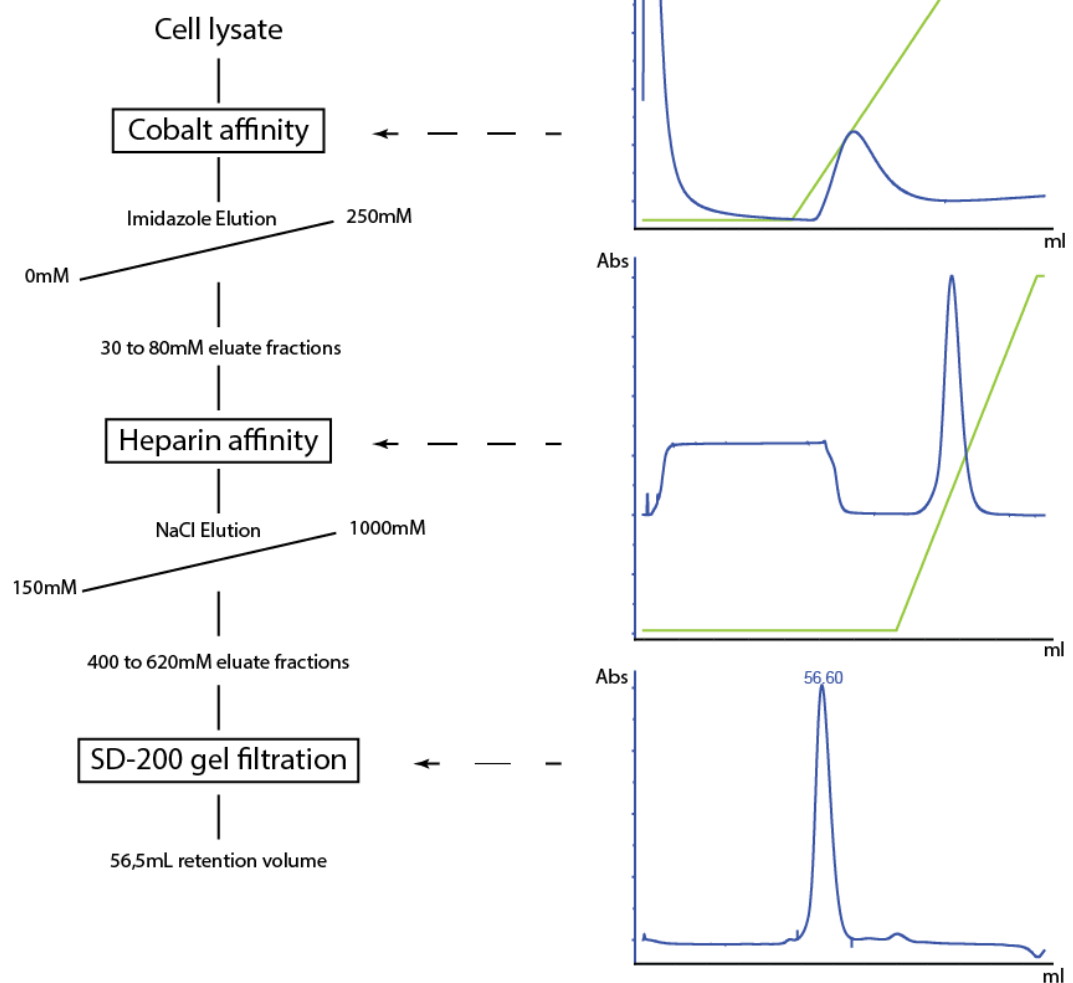


Figure 20 : Diagramme type de la purification de His-Rna15p Rna14p. Pour chaque étape, la courbe en vert représente le gradient d'élution, la courbe en bleue l'absorbance à 280 nm. Le volume indiqué à l'étape de filtration sur gel est le volume d'élution du pic.

Le complexe His-Rna15p Rna14p a exclusivement été produit par coexpression polycistronique durant mon doctorat. Seule la sous-unité Rna15p est en fusion avec l'étiquette 6His. Les deux premières étapes de purification sont menées de la même manière que pour le monomère His-Rna15p, si ce n'est que les fractions d'élutions sont contrôlées par une migration sur un gel SDS-PAGE à 12%. Après l'héparine, les fractions contenant l'échantillon de protéines sont concentrées et chargées sur une colonne Superdex 200. L'élution se fait tampon Tris 25mM pH 7,5, NaCl 250mM et DTT 1mM, à 0,5 ml/min, et un complexe unique est élué à environ 56ml.

Clp1/Pcf1 1 purification Flowchart

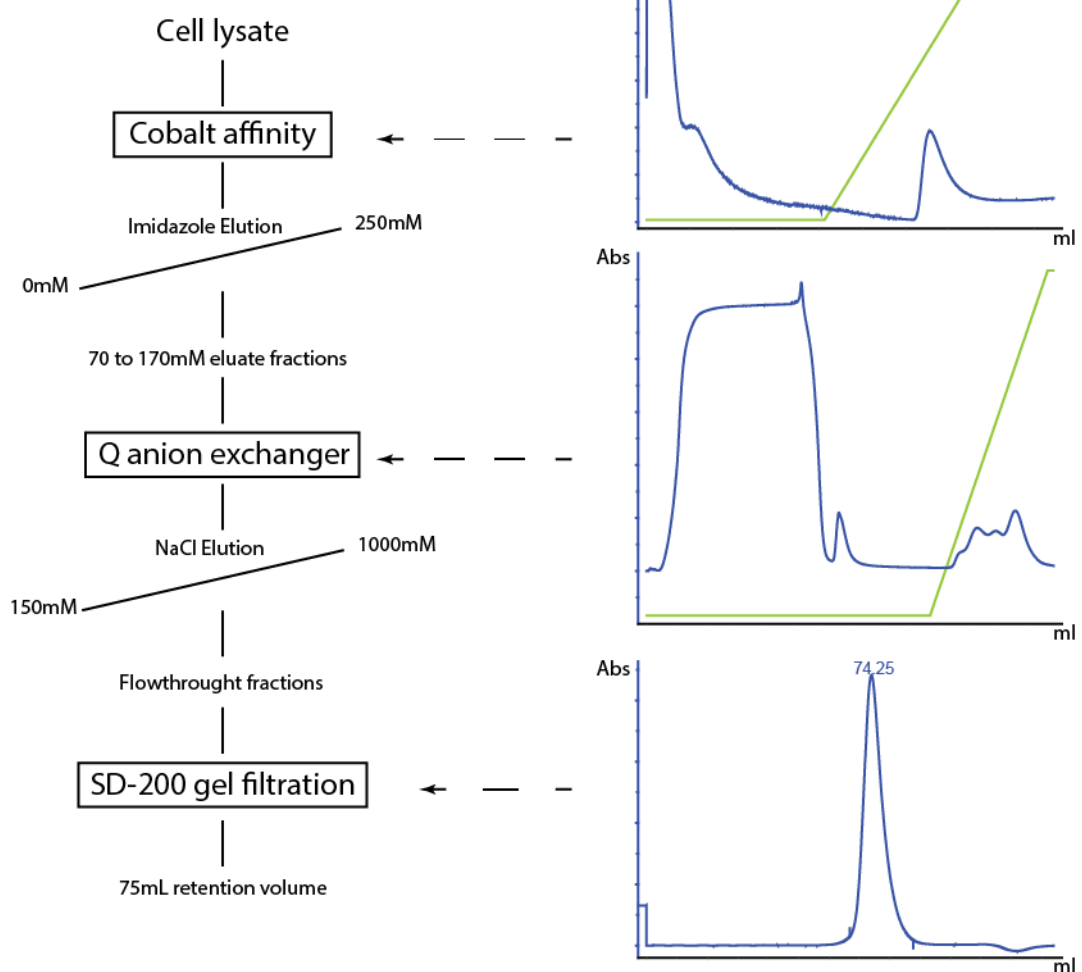


Figure 21 : Diagramme type de la purification de His-Pcf11p (454-563) Clp1p. Pour chaque étape, la courbe en vert représente le gradient d'éluion, la courbe en bleue l'absorbance à 280 nm. Le volume indiqué à l'étape de filtration sur gel est le volume d'éluion du pic.

Le dimère His-Pcf11p Clp1p, que je produis le plus fréquemment sous la forme His-Pcf11p (454-563) Clp1p, n'est pas exprimé à partir d'un vecteur polycistronique à cause des différentes constructions de Pcf11p utilisées. La purification commence par une étape d'affinité contre l'étiquette fusionnée à Pcf11p, dans un tampon Tris 25mM pH 7,5 avec 150mM de NaCl et une éluion par un gradient d'imidazole. Les protéines contenues dans le pic seront contrôlées par migration sur un gel SDS-PAGE 15% et chargées sur une colonne d'ammonium quaternaire échangeuse d'anions. Une éluion avec un gradient de sels permettra d'éliminer la majorité des contaminants et de retrouver les protéines d'intérêt dans le matériel non retenu par les cations. Nous pourrons alors concentrer ces fractions pour les

charger sur une colonne de résine Superdex 75. Une élution avec un tampon Tris 25mM pH 7,5, 250mM de NaCl et 1mM de DTT permettra d'obtenir un pic unique contenant les deux protéines à 75ml d'élution.

CF IA purification Flowchart

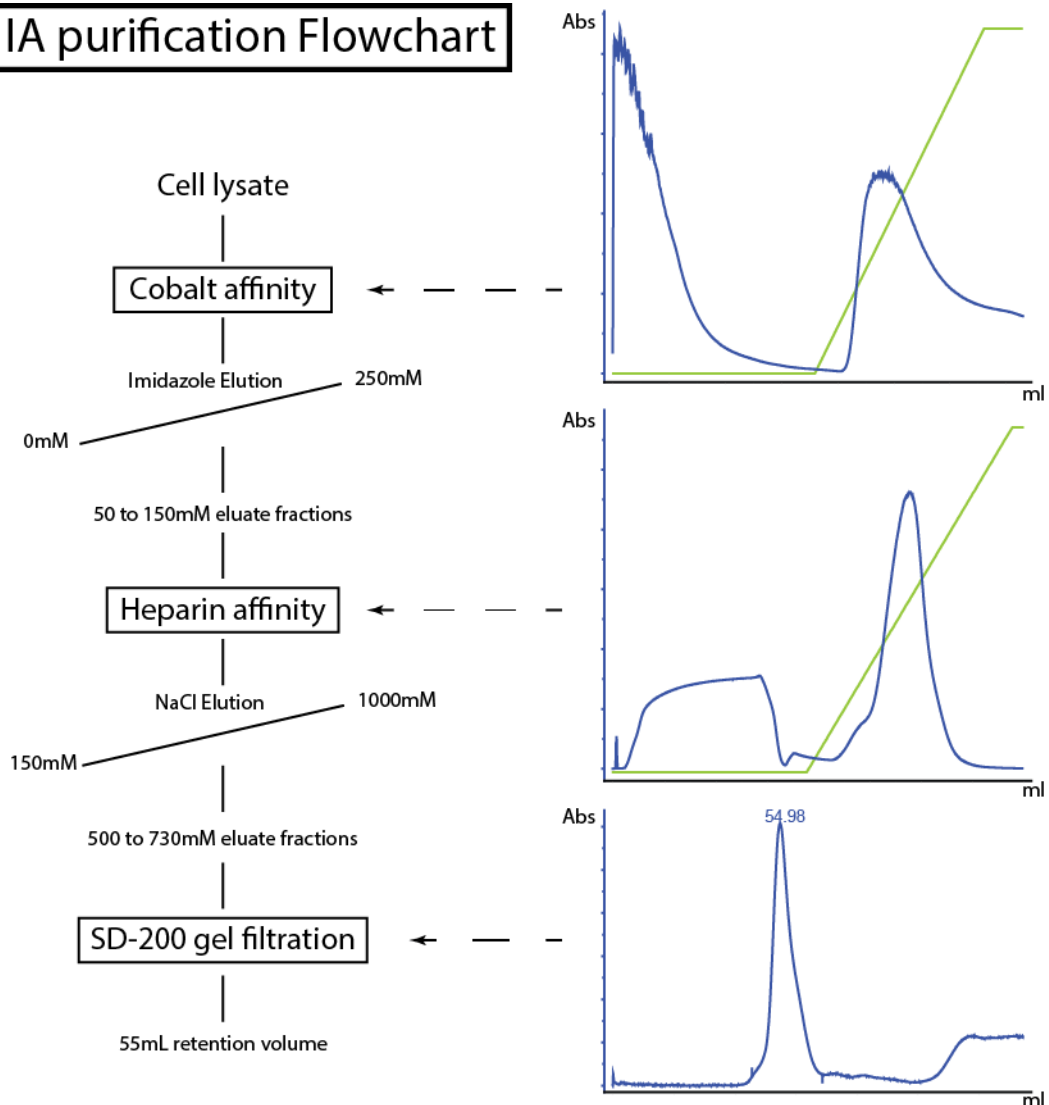


Figure 22 : Diagramme type de la purification de CF IA. Pour chaque étape, la courbe en vert représente le gradient d'élution, la courbe en bleue l'absorbance à 280 nm. Le volume indiqué à l'étape de filtration sur gel est le volume d'élution du pic.

L'expression du complexe entier est le plus souvent effectuée à partir d'un vecteur polycistronique contenant les séquences codantes de Rna14p, Rna15p et Clp1p entières cotransformé avec un vecteur contenant la construction d'intérêt de Pcf11p en fusion avec l'étiquette. L'expression est également régulièrement faite en co-transformant un vecteur polycistronique Rna14p-Rna15p avec un vecteur codant pour Clp1p et un troisième codant pour Pcf11p. Ceci permet à l'étiquette d'être portée par Clp1p tout en gardant la liberté de

modifier la construction de Pcf11p. Le protocole de purification est similaire à celui adopté pour le complexe Rna15p Rna14p. L'étape d'affinité contre l'étiquette est suivie par une colonne Héparine, à laquelle le complexe va se lier. Un épaulement plus ou moins présent peut contenir uniquement Rna15p et Rna14p à cette étape. Une migration des fractions sur un gel SDS-PAGE 12% permettra alors de sélectionner uniquement les plus pertinentes. Après concentration, la purification sera finalisée par une étape d'exclusion de taille sur une colonne de Superdex 200. Un pic unique apparaîtra à l'élution, à un volume d'élution de 55ml, contenant les quatre protéines. Dans certaines conditions, notamment si la purification est effectuée sur plusieurs jours, un épaulement contenant Rna14 et Rna15 viendra se superposer à la fin du pic contenant le complexe entier, et les fractions contenant ces protéines en excès seront écartées.

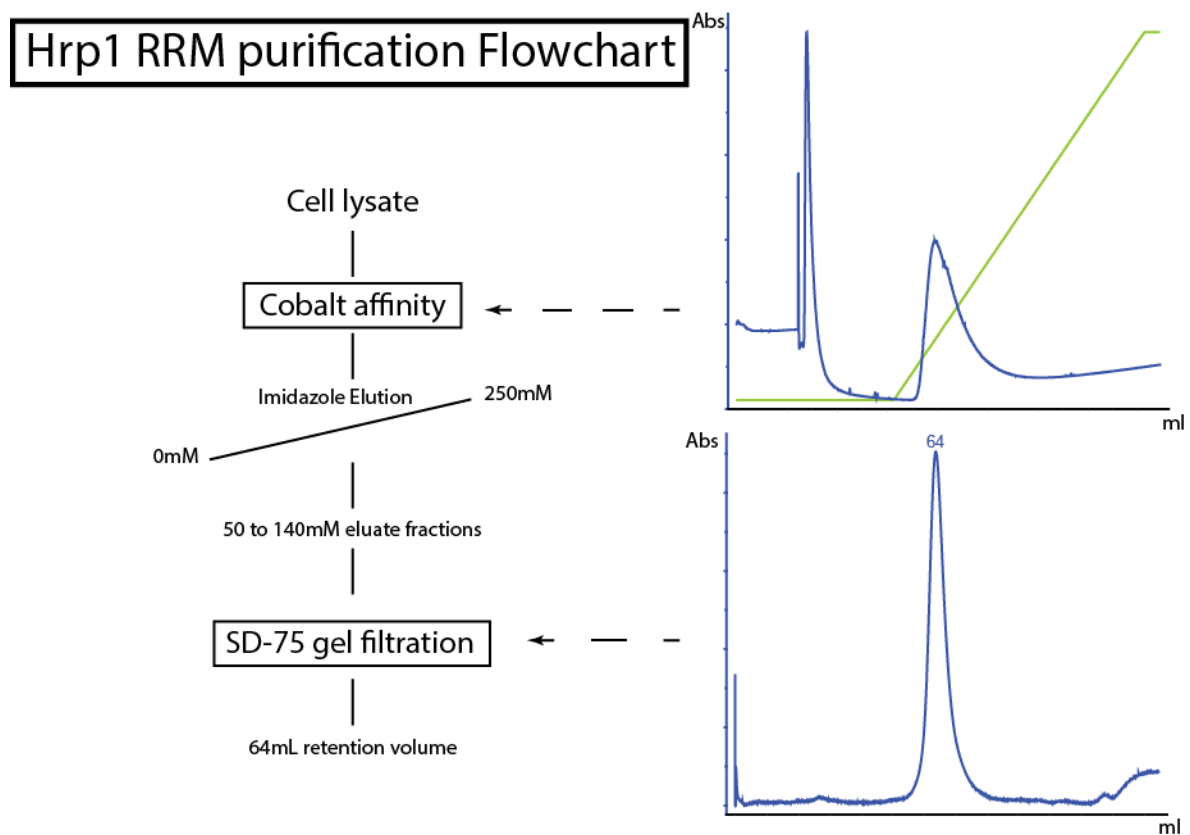


Figure 23 : Diagramme type de la purification de His-Hrp1 (156-322). Pour chaque étape, la courbe en vert représente le gradient d'élution, la courbe en bleue l'absorbance à 280 nm. Le volume indiqué à l'étape de filtration sur gel est le volume d'élution du pic.

La construction contenant les deux RRM de Hrp1p peut être purifiée très simplement, en deux étapes, comme décrit par Perez-Canadillas et al. (Perez-Canadillas 2006, Leeper, Qu et al. 2010). Nous avons juste modifié légèrement les tampons pour les adapter à ceux utilisés

pour CF IA. L'affinité sera ainsi effectuée en Tris 25mM pH 7,5 avec 150mM de NaCl et l'exclusion de taille sera éluée avec un tampon isocratique Tris 25mM pH 7,5, NaCl 250mM et DTT 1mM. A chaque étape, la pureté est vérifiée par une migration sur un gel SDS-PAGE 15%.

3.2.3. Cristallisation

3.2.3.1. Principes

La cristallisation est le processus entraînant la formation de cristaux solides à partir d'un soluté. Le principe de base de la cristallisation est de modifier progressivement les conditions de solubilité dans la solution (température, concentration) pour entraîner une saturation et une nucléation des molécules de soluté. La nucléation, ou germination, est une organisation de la matière selon la maille d'une structure cristalline, formant alors un noyau de quelques molécules qui pourra être le germe et le support de la croissance cristalline. Cependant, dans un état de saturation, la protéine peut également précipiter de manière non organisée et former des agrégats. La capacité d'une protéine à cristalliser dépendra en fait directement de l'état le plus énergétiquement favorable, entre le précipité et le noyau, dans les conditions données (Figure 24).

La recherche de ces conditions de nucléation est un processus empirique consistant à tester différentes conditions (concentration initiale, température, taille des gouttes) et différents additifs (sels, pH, précipitants). Au laboratoire, nous faisons des essais de cristallogénèse par diffusion de vapeur, en goutte assise dans la plupart des cas (Figure 25).

Dans cette installation expérimentale, un volume identique de protéine et de solution de cristallisation sont mélangés pour former la goutte de cristallisation déposée sur un piédestal. Le réservoir adjacent est rempli de la même solution de cristallisation, de manière à ce que le volume de la goutte soit négligeable en regard de celui du réservoir. Dans l'espace alors scellé nous aurons donc deux volumes à des concentrations différentes, ce qui va entraîner une diffusion de l'eau du milieu le moins concentré au milieu le plus concentré. Grâce à ce phénomène, la concentration en protéine et en précipitant va augmenter progressivement dans la goutte (Figure 26, chemin A)

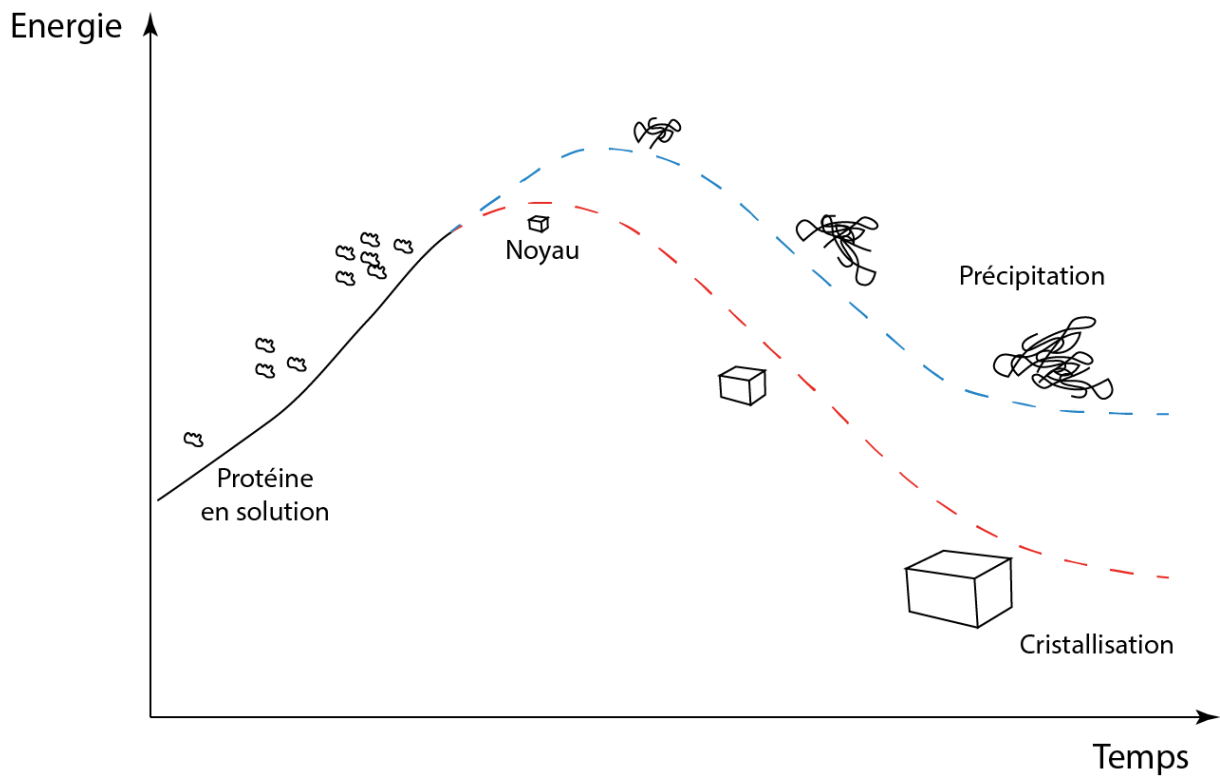


Figure 24 : Diagramme d'énergie pour la cristallisation. Dans l'exemple illustré ici, l'état avec l'arrangement cristallin (le noyau) à une énergie libre plus basse que celle des arrangements irréguliers (précipitation), et sera donc favorisé. Cette configuration est donc très favorable à la cristallisation.

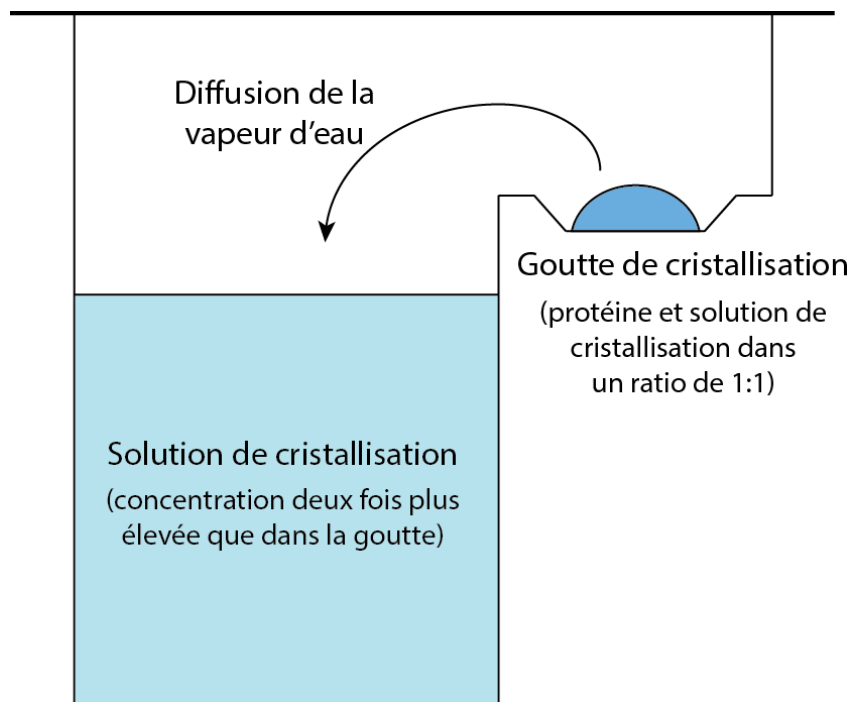


Figure 25 : Stratégie de diffusion de vapeur en goutte assise.

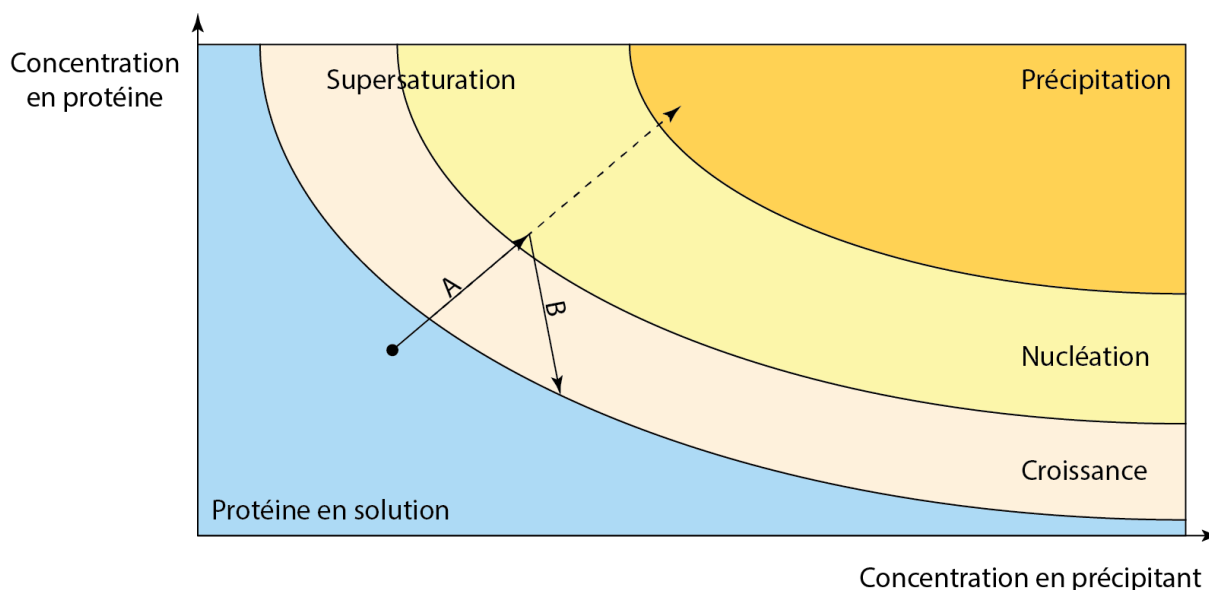


Figure 26 : Diagramme des phases lors de la cristallisation protéique. Le chemin A + B représente un exemple d'évolution des conditions lors de la cristallisation dans un système par diffusion de vapeur.

Il n'existe pas de règles précises concernant le comportement des protéines lors de la cristallogénèse, mais quatre zones peuvent être déterminées en fonction des concentrations respectives en protéine et en agent précipitant. La Figure 26 présente le chemin théorique de l'évolution de ces concentrations lors d'une cristallisation par diffusion de vapeur. Les molécules présentes dans la goutte vont se concentrer durant la diffusion de vapeur (chemin A), entraînant une supersaturation. Cet état de supersaturation est propice à la nucléation et à la précipitation, et l'état le plus favorable énergiquement aura plus de chance de se produire. Une nucléation entraînera une baisse de la concentration en protéine (chemin B) et une croissance du cristal jusqu'à un équilibre.

3.2.3.2. Conditions testées

Pour augmenter les probabilités de parvenir à cristalliser une protéine, il faut balayer de manière pertinente mais exhaustive différentes conditions de cristallisation. Le Tableau 1 résume les conditions que nous avons explorées pour les différents projets.

Projet	Constructions testées	Conditions testées	Concentrations testées (mg/ml)	T° testées	Hits
Domaine N-terminal de Pcf11p	His-Pcf11pkl (1-234)	ION-PEG-NATRIX Wizard ProPlex I&II MbClass MbClass II	4 8 30	20°C	

Projet	Constructions testées	Conditions testées	Concentrations testées (mg/ml)	T° testées	Hits
		Index Protein Complex Anions Cations Crystal Classics PEGs PEGs II JCSG+ Morpheus Salt Rx			
Organisation de CF IA	CF IA His-Pcf11p	Index Protein Complex MbClass Classics Screen NaCl Screen Mg formate Screen MPD Screen MgCl ₂ Screen pH Na Citrate Screen pH Hepes	1 1,1 1,6	20°C	
	CF IA His-Pcf11p (298-563)	Index	5,5	15°C	
	CF IA His-Pcf11p (331-563)	MbClass MbClass II MIDAS Protein Complex PACT ProPlex I&II PEGs PEGs II Salt Rx Wizard I&II Index Screen PEG 3000 Screen PEG 3350 Screen PEG 4000 Screen PEG 5000 Screen pH Hepes	6 7,9	4°C 15°C 20°C	
	CF IA His-Pcf11p (331-626)	Screen PEG 2000 Screen PEG 3000 Screen PEG 3350	4 9	20°C	

Projet	Constructions testées	Conditions testées	Concentrations testées (mg/ml)	T° testées	Hits
		Screen PEG 4000 Screen PEG 5000 Screen pH Hepes			
	CF IA His-Pcf11p (298-626)	JCSG+ Crystal Cryo Cation MIDAS Morpheus Protein Complex Proplex Classics PACT PEGs	6,5 7	15°C	
	CF IA His-Rna15p, Pcf11p (331-563)	Morpheus Wizard I&II MbClass II PACT PEGs PEGs II MIDAS Index Protein Complex Anions Cations	11,2	20°C	
	CF IA His-Clp1p, Pcf11p (298-626)	Wizard I&II Index Cations Anions Classics	5	20°C	
	CF IA His-Pcf11kl (322-518)	Index ProComplex MbClass MbClass II Cations Anions Proplex Classic Additive Wizard III / PEG ion JCSG+ Morpheus MIDAS Cryos PEGs	1 9,2 10 11 14 30 42	4°C 15°C 20°C	oui

Projet	Constructions testées	Conditions testées	Concentrations testées (mg/ml)	T° testées	Hits
		PEGs II Salt Rx Crystal Protein Complex PACT Screen PEG 550 Screen PEG 2000 Screen PEG 3350 Screen PEG 4000 Screen PEG 5000 Screen NaCl Screen NH ₄ Cl Screen KCl Screen Mg formate Screen MPD Screen pH Na Citrate Screen pH Hepes Screen pH Tris Screen pH TBP			

Tableau 1 : Récapitulatif des différents essais de cristallogénèse effectués durant mon doctorat.

3.2.3.3. Cristaux obtenus

Nous résumerons dans le Tableau 2 les conditions de cristallogénèse notables identifiées durant le screening des différentes constructions et conditions.

Protéine	Date	Screen Commercial	Conditions initiales du positif	Amélioration et diffraction
CF IA His-Pcf11p (331-563)	19/01/2012	MbClass II E3	Protéine à 6 g.l 12% PEG 4000 HEPES 0,1M pH 7,5	résorption
		MbClass II B3	Protéine à 6 g.l 0,1M HEPES pH 7,5 1,5M LiSO ₄	Barettes Diffraction de sels
CF IA His-Pcf11kl (322-518)	21/01/2013	PACT E8, E11, F9, H9	Protéine à 10 g.l 20% PEG 3350 Ions Na	Amélioration, collecte à 3 Å
		Cations F9	Protéine à 10 g.l 0,1M Tris-HCl pH 8,5 3,2M NaCl	

		Salt Rx A7	Protéine à 10 g.l 3,5M NH ₄ Cl 0,1M BTP pH 7,0	Amélioration, acquisition à 3,5-4 Å
		Salt Rx B2	Protéine à 10 g.l 3,2M NaCl 0,1M Tris pH 8,5	
	03/03/2014	Mb Class E5	Protéine à 9,2 g.l 0,1M NaCl 0,1M Tris pH 8,5 12% PEG 4000	aiguilles
		Classics H12	Protéine à 9,2 g.l 0,1 MES pH 6,5 12% PEG 2000	Plaquettes Amélioration, acquisition à 3 Å
		Proplex B4	Protéine à 9,2 g.l 0,2M NaCl 0,1M MES pH 6,5 10% PEG 4000	

Tableau 2 : Récapitulatifs des conditions de cristallogénèse commerciales pour lesquelles des objets cristallins ont pu être observés.

Les premiers positifs que nous avons obtenus sur CF IA, lors des essais sur CF IA His Pcf11p (331-563), ont été particulièrement frustrants car l'un s'est avéré être un cristal de sels tandis que l'autre s'est simplement résorbé en 48h. Ce comportement est à priori de bon augure quant à la nature du cristal, car les cristaux de sels sont très stables, ce qui signifie qu'il y a de bonnes chances pour que ce soit de la protéine qui ait cristallisé. Malheureusement, malgré de nombreuses tentatives, je n'ai pas été capable de reproduire ou au moins de réobserver de cristallisation dans ces conditions.

Néanmoins, un an plus tard, l'homologue de *K. lactis* a permis l'obtention de cristaux dans plusieurs conditions différentes (Figure 27, B). Après optimisation des conditions, nous avons réussi à passer de bouquets d'aiguilles à des barrettes plus massives. Des acquisitions ont pu être faites sur les lignes de lumière ID23, à l'ESRF et PX1 à SOLEIL. Les meilleures acquisitions ont des tâches à partir de 3,5 Å. L'indexation est faite avec XDS (Kabsch 2010), révélant une maille d'ordre hexagonal $P6(3)22$. En considérant la symétrie très élevée de ce système, reflétée par la redondance de 8,6 (Tableau 3) et le faible nombre de réflexions uniques enregistrées, nous avons estimé avec l'aide de P. Legrand (ligne PX1, SOLEIL) qu'environ 400 résidus au grand maximum pouvaient être contenus dans l'unité asymétrique. Nous avons alors testé d'effectuer des remplacements moléculaire avec des fenêtres

d'environ 250 résidus parmi les structures déjà connues au sein du CF IA, sans que rien ne donne de résultats. J'ai également fait une recherche dans la PDB avec les paramètres de mailles, dans l'hypothèse que cela soit un contaminant bactérien récurrent, mais rien n'est sorti. Afin d'identifier directement le contenu du cristal, j'ai tenté d'en collecter un nombre suffisant afin de les déposer sur gel et peut être pouvoir identifier un bande claire, mais cela n'a rien donné. J'ai alors fait quelques essais d'incorporation de sélénométhionine lors de l'expression du complexe, avec pour objectif de d'exploiter les signaux de diffraction anormale, mais l'expression s'est retrouvée perturbée. En considérant tout cela, nous avons décidé de ne pas insister sur ce fragment pour continuer à se concentrer sur le complexe entier.

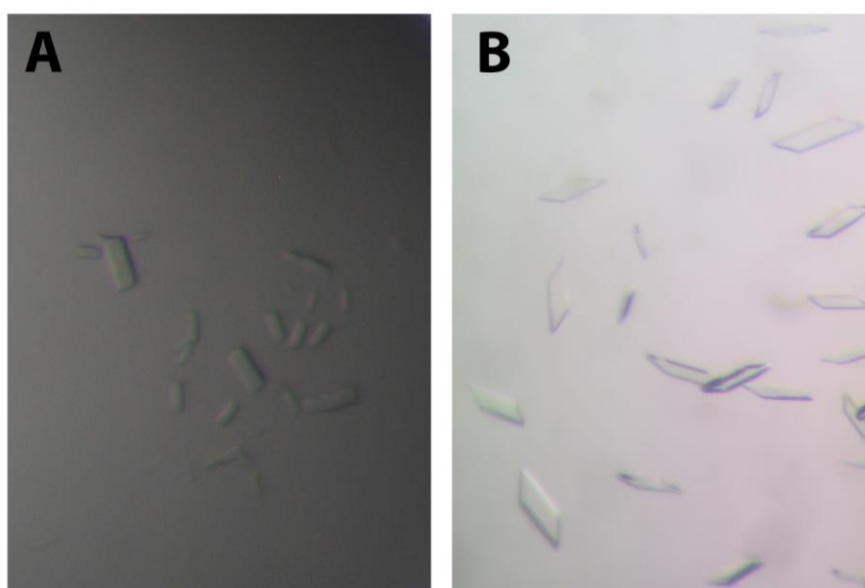


Figure 27 : Exemples de cristaux obtenus après optimisation des conditions de cristallogénèse. **A.** Cristaux de CF IA His-Pcf11kl (322-518) concentrée à 42g.l obtenus après 6 jours à 20°C en 0,1M BTP pH9, 2M NH₄Cl et 2% PEG 2000MME. **B.** Cristaux de CF IA His-Pcf11kl (322-518) concentrée à 9,7g.l obtenus en 16% PEG 2000 et 0,1M MES pH 6.

A force de persévérance et d'amélioration de la purification, de nouveaux essais avec des solutions commerciales déjà testées permirent d'identifier de nouvelles conditions positives. D'aspect initial relativement identique (botte d'aiguilles), l'optimisation conduit vers une forme cristalline légèrement différente (Figure 27). Des acquisitions de bonne qualité (tâches visibles jusqu'à 2,5Å) ont été effectuées, et une indexation nous permet d'obtenir des paramètres de mailles : $a=118$, $b=60$, $c=123,9$, $\beta=101,6$ pour un groupe d'espace qui semble être C2. Une rapide recherche dans la Protein Data Bank (www.rcsb.org/pdb) avec ces paramètres de mailles nous permet de sortir une structure unique référencée sous PDB ID=4E6H. Il s'agit de la structure des domaines HAT de Rna14kl publiée par Paulson (Paulson

and Tong 2012). Résolue à 2,3Å, elle est constituée par Rna14kl (28-582) en complexe avec lui-même, comme cela avait déjà pu être décrit chez d'autres espèces (Bai, Auperin et al. 2007, Legrand, Pinaud et al. 2007).

3.2.3.4. Structure de Rna14kl

Une fois les images indexées, intégrées et mises à l'échelle par XDS et XSCALE, XDSCONV (tous trois présents dans le XDS Package (Kabsch 2010)) est utilisé pour générer un fichier contenant les intensités dans un format exploitable par le programme Phaser (McCoy, Grosse-Kunstleve et al. 2007) dans la suite Phenix (Adams, Afonine et al. 2010). Suspectant très fortement que nos cristaux sont également constitués de Rna14kl partiellement dégradé, j'ai en effet décidé de tenter un remplacement moléculaire (McCoy 2007) avec la structure de Paulson (PDB ID=4E6H). Cette stratégie a été très efficace, car un premier modèle a ainsi pu être proposé, avec un RMSD=0,51Å. Des étapes d'affinement de ce modèle ont alors été réalisées avec Buster ou phenix.refine, tandis que la construction est effectuée avec Coot (Emsley, Lohkamp et al. 2010). Les paramètres finaux sont présentés dans le Tableau 3, seconde colonne « Cristaux obtenus en 2014 ».

La Figure 28 présente le modèle final déposé dans la PDB sous l'ID numéro XXXX (Je compte déposer cette structure en rendant ma thèse. Cela sera donc complété dans la version finale). Dans le panneau A, nous pouvons voir que la structure que j'ai résolue est extrêmement proche de celle résolue par Paulson. Le RMSD entre les deux structures est en effet uniquement de 0,348Å. Si l'on examine la structure, cette différence provient essentiellement des boucles, les hélices alpha ayant une position très bien conservée. Il faut également remarquer les quelques acides aminés que je suis parvenu à construire mais qui sont absents de la structure publiée. En N-terminal, 7 résidus ont ainsi été ajoutés au modèle (B), notamment grâce à une construction différente de la boucle formée par les résidus 32 à 28.

	Cristaux obtenus en 2013	Cristaux obtenus en 2014	Structure de Paulson et al.
PDB ID			4E6H
Statistiques cristallines			
Groupe d'espace	<i>P</i> 6 (3) 22	<i>C</i> 2	<i>C</i> 2
a, b, c (Å)	105,53 105,53 127,59	117,83 60,26 123,58	117,4 59,8 123,5
α, β, γ (°)	90,0 90,0 120,0	90,0 101,988 90,0	- 101,6 -
Source du rayonnement	ID23 (ESRF)	PX1 (SOLEIL)	X29A (NSLS)
Longueur d'onde (Å)	0,8729	0,9801	-
Résolution (Å)	48,73 – 3,49	48,18 – 2,38	50 – 2,3
Réflexions totales	49 486 (7 730)	149 101 (23 110)	155 841
Réflexions uniques	5 726 (873)	33 273 (5 374)	-
Redondance	8,6	4,5	4,3
Complétude (%)	99,7 (98,6)	99,5 (97,6)	89 (68)
Rsym (%)	27,07 (91,6)	6,40 (56,7)	5,3 (24,7)
I/ σ	9,59 (2,74)	14,60 (2,35)	21,7 (4,2)
Statistiques d'affinement			
Nombre de résidus		563	553
Molécules d'eau		113	230
Rwork (%)		0,2086	0,210
Rfree (%)		0,2417	0,270
$\phi\psi$ Favorisé (%)		95,72	91,6
$\phi\psi$ Autorisés (%)		100	97,6

Tableau 3 : Résumé des informations de l'acquisition cristallographique et du traitement des images. Les cristaux présentés correspondent aux meilleures acquisitions effectuées sur un ensemble de cristaux. Les nombres entre parenthèse correspondent aux données pour l'enveloppe de plus haute résolution. Les données relatives à la structure de Paulson & al. ont été recensées dans l'article correspondant (Paulson and Tong 2012) ou directement dans le document « .pdb ».

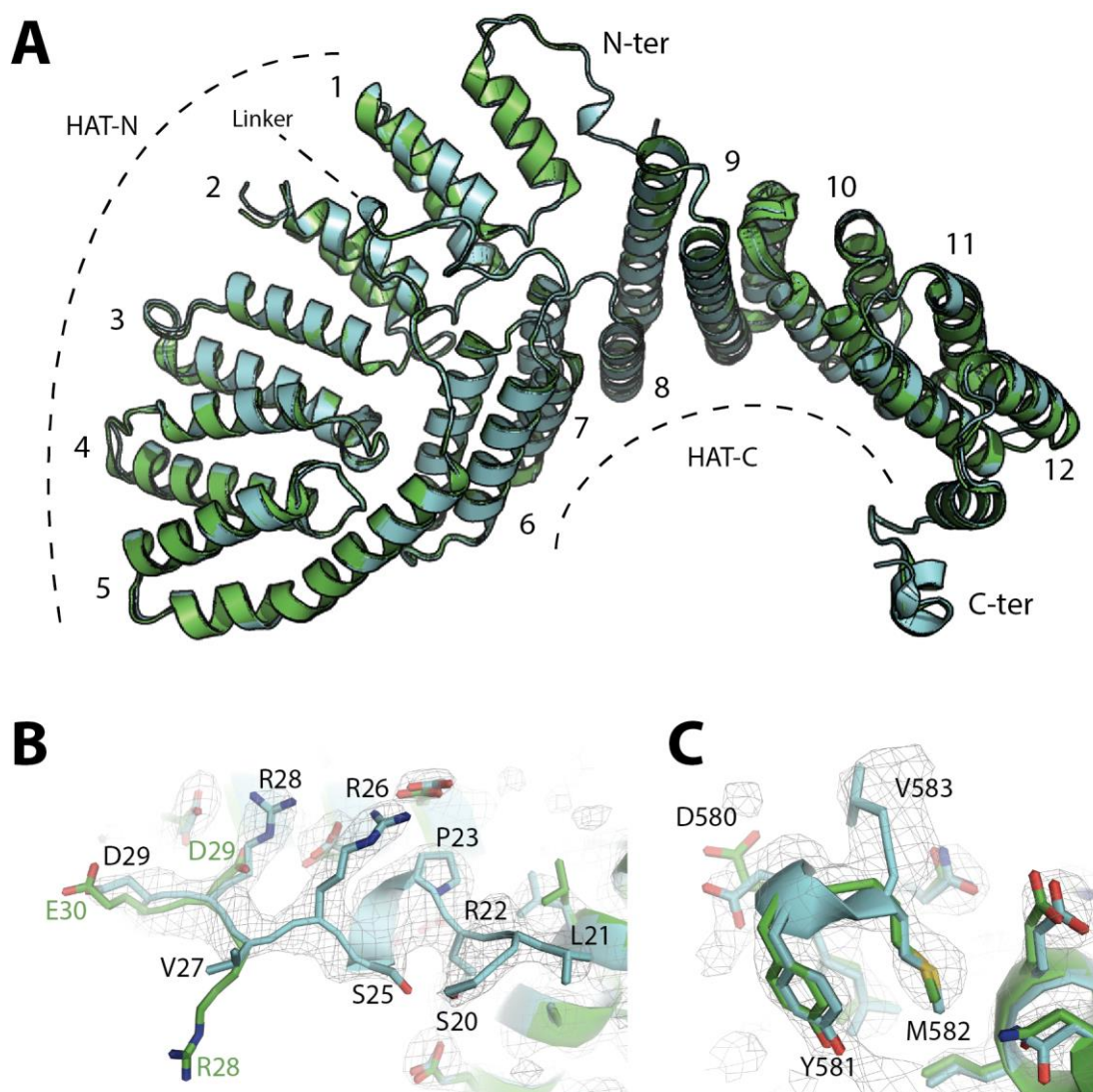


Figure 28 : Structure de Rna14kl (20-583) résolue à 2,38Å (PDB ID=XXXX), représentée en bleue. La structure de Paulson et al. (PDB ID=4E6H) est représentée en vert. **A.** Superposition des deux structures (RMSD = 0,348 Å) **B.** Les résidus N-terminaux du modèle final, placé dans la carte 2Fo-Fc (gris), illustrent la différence de construction des résidus 32 à 28 et les résidus 27 à 21 supplémentaires. **C.** Le C-terminal du modèle final, placé dans la carte 2Fo-Fc (gris), illustrant le résidu supplémentaire.

Il est intéressant de remarquer que malgré une résolution légèrement inférieure (2,3Å vs 2,38Å), les statistiques d'affinement de ma structure sont globalement meilleures. Dix résidus supplémentaires ont pu être attribués, les Rwork et Rfree sont plus faibles et les résidus sont mieux positionnés dans le diagramme de Ramachandran (Tableau 3). Cela peut probablement être expliqué par le fait que le jeu de données à partir duquel j'ai travaillé était de meilleure qualité. Si l'on regarde les données cristallographiques, on doit en effet remarquer que mon jeu était plus complet, avec 99,5% (97,6%) de complétude contre 89% (68%). La différence est particulièrement marquée à très haute résolution : mon jeu est quasiment complet là où 22% de l'information est manquante chez Paulson et al. Malgré cette

différence en complétude, les deux jeux de données ont à peu près la même redondance (4,5 vs 4,3). Par contre, le rapport signal/bruit (I/σ) est très élevé sur ses images [14,60 (2,35) vs 21,7 (4,2)], complexifiant leur analyse. Il faut également remarquer la différence dans le nombre de molécules d'eau attribuées (113 contre 230). Malgré cela, les facteurs R/Rfree de Paulson sont restés relativement hauts, ce qui pourrait indiquer la présence de molécules d'eau là où j'ai réussi à ajouter des résidus supplémentaires.

Il est intéressant de noter que les deux cristaux sont issus de la dégradation *in vitro*. En effet, Paulson & al. cherchaient à identifier des conditions pour résoudre le complexe Rna14kl-Rna15kl, tandis que je m'attachait à cristalliser le CF IA kl en entier. Dans le cas de Paulson & al., le produit de dégradation a cristallisé en 50mM Hepes (pH 7.0), 12% (w/v) PEG3350, et 1% (w/v) tryptone. Au laboratoire, les meilleures diffractions ont été obtenues sur des cristaux poussés en 0,1M BTP (pH 9), 3M NH₄Cl, 8% (w/v) Glycérol. Cette différence est certainement à l'origine d'une partie de la différence dans la qualité des jeux de données, la ligne de lumière, les différentes manipulations subies par les cristaux et les conditions de cryogénisation entrant aussi en ligne de compte.

3.2.3.5. Conclusions

La cristallogenèse est une étape empirique, à laquelle il convient de consacrer du temps et de l'énergie, sans se décourager. Si nous essayons de dresser un bilan de notre expérience avec les essais de cristallisation du CF IA, nous pouvons raisonnablement dire que nous avons testé un large panel de conditions, criblant concentration, température, pH, agent précipitant, sels, additifs divers et séquence protéique. En terme de constructions notamment, où nous avons tenté de limiter au maximum l'impact des régions non structurées. Nous avons également testé différents homologues. Cette option a d'ailleurs permis d'amener les quelques résultats que nous avons pu obtenir. Car si l'on excepte un cristal observé en janvier 2012 qui s'est dissout et que je ne suis pas parvenu à observer à nouveau (Tableau 2), tous les cristaux ont été obtenus avec les séquences codantes de *Kluyveromyces lactis*. Finalement, aucun ne s'est avéré être constitué par le complexe entier, au mieux les domaines HAT de Rna14kl.

Il semble donc que CF IA soit particulièrement réfractaire à la cristallisation. Sa mobilité intrinsèque, amenée notamment par le RRM et les nombreux « linkers », constituants

probablement la principale limite. Nous devons donc nous pencher vers des approches qui nous permettront soit de concilier cette contrainte soit de la limiter au maximum. Il sera également pertinent de s'attacher à « raboter » le complexe un peu plus. Le RRM de Rna15p notamment, peut encore être retiré sans limiter la formation du CF IA.

3.2.4. Approche dynamique par la diffusion aux petits angles

3.2.4.1. Principe

Beaucoup des efforts fournis durant cette thèse se sont concentrés autour de l'approche de SAXS. Ceci a été fait en collaboration avec Javier Perez et Pierre Roblin de la ligne SWING au synchrotron SOLEIL.

Si nous avons consacré beaucoup de nos efforts à acquérir et exploiter des données de SAXS (Small angle X-ray scattering), c'est car c'est une technique de biologie structurale complémentaire de toutes les autres, grâce laquelle nous espérons pouvoir combiner toutes les données déjà connues. Le SAXS est en effet une technique qui fournit des informations structurales sur des macromolécules en solution, ce qui est son principal avantage, et il est très rapide de faire des acquisitions. Cependant, l'information obtenue est une information moyennée sur la taille et l'enveloppe des objets présents dans la solution car le rayonnement X va illuminer les macromolécules orientées aléatoirement dans la solution. Le SAXS ne distinguera donc pas différents objets d'une solution hétérogène, et ne sera pas non plus capable de mesurer la position des atomes dans la molécule.

Cependant, combiné à des données structurales déjà connues, les données de SAXS peuvent permettre de construire un modèle global dont la diffusion théorique correspond à la diffusion expérimentale.

Pour pallier au maximum à la limite qui nous oblige à travailler avec des informations de diffusion correspondant aux contributions de toutes les molécules présentes dans la solution, J. Perez et P. Roblin ont mis au point une innovation technique sur la ligne SWING en montant un système HPLC en amont de l'acquisition des données. Très performant, ce montage permet de faire l'acquisition de données sur l'échantillon directement en sortie de colonne d'exclusion sphérique et ainsi d'homogénéiser au maximum la population contribuant à la diffusion. La Figure 29 illustre le montage expérimental (A) et la procédure de

sélection des plages d'acquisition des données. Lors d'une première injection de calibration, la lecture de l'absorbance en sortie de colonne nous permet de définir un créneau d'acquisition de la contribution du tampon et un créneau pour l'acquisition de la protéine d'intérêt (B). Une seconde injection servira à faire les acquisitions de données le long du pic, donnant un profil concentration-dépendant (C). Une transformée de fourrier va alors nous permettre de visualiser très simplement le rayon de giration calculé pour les molécules contribuant à la diffusion. Nous sélectionnerons alors exclusivement les images sous le pic pour lesquelles le rayon de giration est constant. Tout ceci est effectué à l'aide du logiciel FOXTROT, développé par l'équipe de SWING. Les traitements de données suivants seront réalisés avec la suite logicielle ATSAS pour le traitement des données de diffusion aux petits angles.

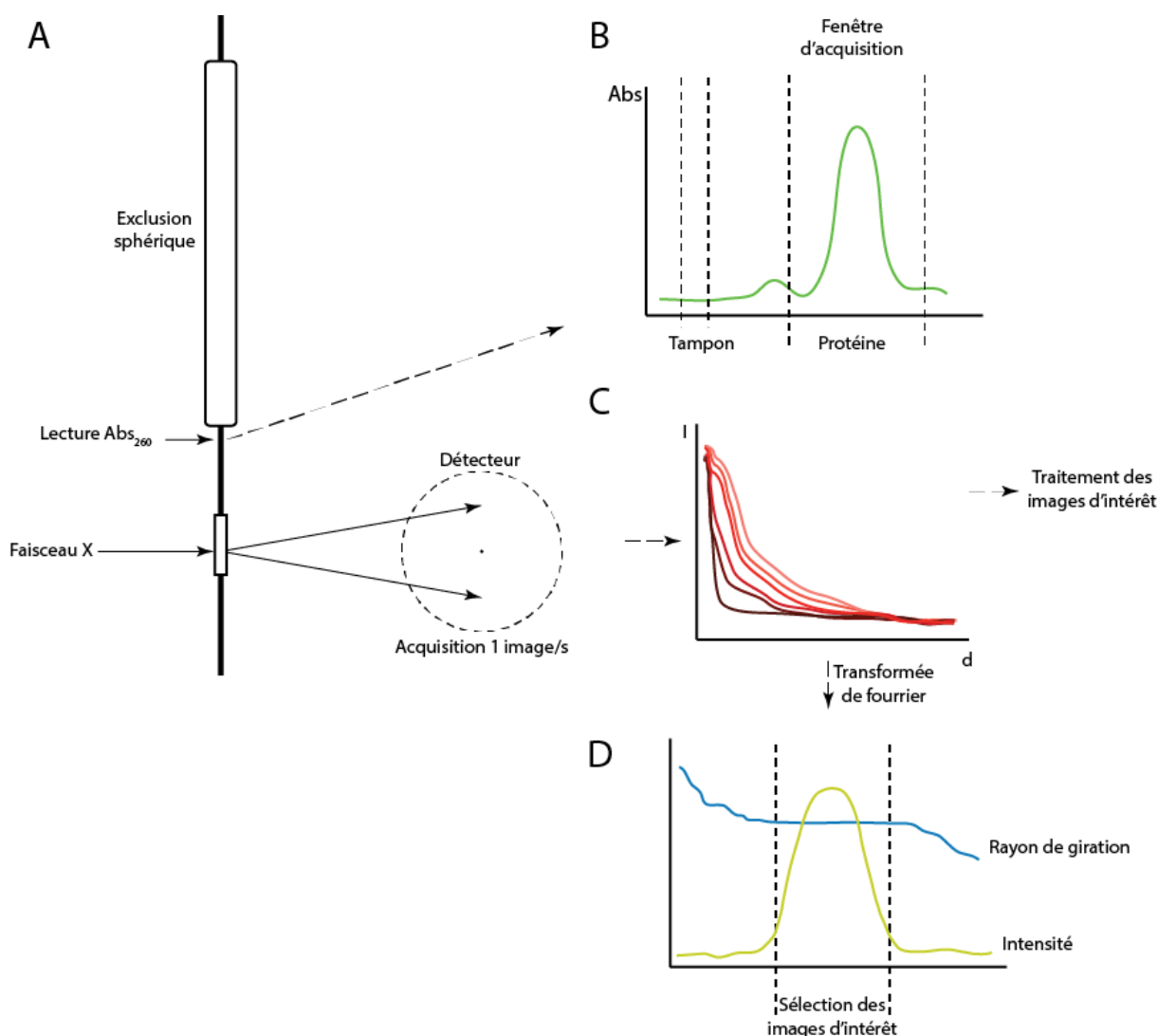


Figure 29 : Montage expérimental spécifique de la ligne de lumière SWING à SOLEIL.

L'historique des analyses de CF IA par la technique de SAXS est relativement pauvre. En fait, seuls Gordon et al. (Gordon, Shikov et al. 2011) utilisent cette technique pour comparer le dimère de CstF-77 à l'hétérodimère Rna14p-Rna15p. Ils ne font état en fait d'aucune tentative de modélisation, se contentant de calculer un rayon de giration (Cstf-77 : 70,9 Å ; Rna14p-Rna15p : 67,0 Å) et la probabilité de chaque distance inter atomique ($p(r)$). Comme ils observent que les petites distances ($r < 50$ Å) sont plus probables chez Rna14p-Rna15p, ils en déduisent que les deux copies de Rna15p se situent probablement proches du centre du dimère formé par Rna14p.

Il apparait donc qu'il existe avec le SAXS une piste non explorée de l'analyse de CF IA. Le SAXS ayant besoin d'être alimenté par des données structurales pour être exploité de manière optimale, nous estimons avoir maintenant suffisamment de connaissances pour qu'il soit pertinent de consacrer des efforts à cette approche.

3.2.4.2. Echantillons testés

Dans la logique de modélisation de CF IA, j'ai axé mes constructions sur les fragments dont la structure est déjà résolue. Si le SAXS peut en effet positionner des éléments connus entre eux pour produire un modèle avec une diffusion théorique proche de la diffusion expérimentale, il est extrêmement affecté dans les modélisations par la liberté générée par les régions inconnues qu'il convient donc de minimiser autant que possible.

- HisRna15p
- HisCstF77ec
- HisRna14p
- HisRna15p-Rna14p
- HisPcf11p(454-563)-Clp1
- HisRna15p-Rna14p-Clp1p
- CF IAsc
 - His-Rna15p, Pcf11p (298-563)
 - His-Rna15p, Pcf11p (298-626)
 - His-Pcf11p (298-626)
 - His-Pcf11p ΔQ
 - His-Clp1p, Pcf11p (298-626)

3.2.4.3. Analyses et essais de modélisation

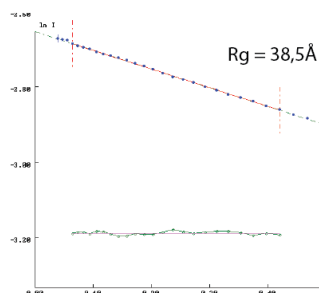
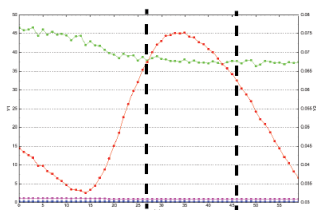
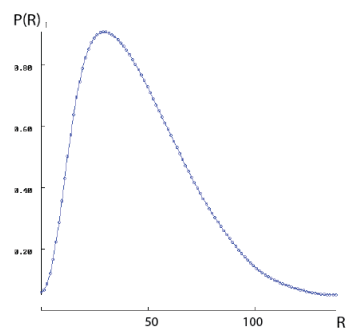
Le programme PRIMUS (Konarev, Volkov et al. 2003), de la suite logicielle ATSAS, est utilisé pour les premières étapes de traitement des données expérimentales. La première étape consiste à sélectionner les clichés situés sous le pic d'absorbance pour lesquels le rayon de giration estimé est constant (Figure 30, premier graphique). Cette étape est au cœur de l'innovation apportée par le couplage avec un système HPLC sur la ligne SWING, car elle nous permet d'exclure totalement de potentiels agrégats et nous assure de travailler avec des clichés de diffusion d'un matériel homogène. PRIMUS va alors nous permettre de moyenner les images sélectionnées, puis de soustraire la contribution du solvant. Nous obtenons alors la courbe de l'intensité de la diffusion en fonction de l'angle de diffusion pour notre protéine d'intérêt. PRIMUS permet alors de calculer une valeur plus précise que l'estimation précédente pour le rayon de giration de notre objet (Figure 30, « R_g », second graphique) via une approximation de Guinier. La courbe de diffusion peut également être exploitée par GNOM (Svergun 1992) afin de calculer une fonction de distribution de taille, la $P(R)$ (Figure 30, troisième graphique). Cette fonction représente la probabilité qu'une distance interatomique, notée R , soit présente dans le système à l'origine de la diffusion. La $P(R)$ doit donc, en tout état de cause, avoir une valeur nulle pour $R=0$, deux atomes ne pouvant se superposer, ce qui fait de cette fonction un bon contrôle de la qualité des données. Nous pouvons ainsi voir que pour les acquisitions A et C, la $P(R)$ ne démarre pas à 0, trahissant des données de qualité moyenne.

Dans notre cas, la courbe de diffusion sera alors exploitée pour tenter de modéliser des enveloppes pour nos constructions ou pour agencer et lier dans l'espace les domaines connus en un modèle dont la diffusion théorique serait cohérente avec la diffusion expérimentale. La modélisation *ab initio* de l'enveloppe est effectuée avec DAMMIF (Franke and Svergun 2009) ou DAMMIN (Svergun 1999) tandis que la modélisation du complexe est effectuée avec CORAL (Petoukhov, Franke et al. 2012), SASREF (Petoukhov and Svergun 2005), DAMAVER (Volkov and Svergun 2003) ou DaDiModO (Evrard, Mareuil et al. 2011).

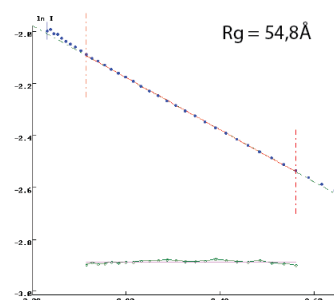
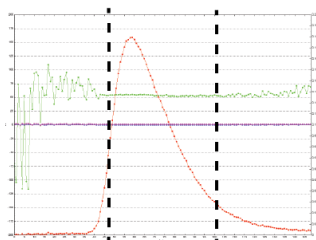
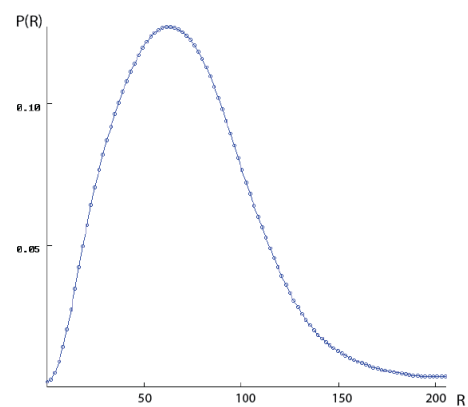
La Figure 31 représente le travail effectué sur le complexe Pcf11p (454-563) Clp1p. Bien que cette structure soit très bien caractérisée, nous pouvons voir dans le panneau A que les différentes modélisations effectuées (avec DaDiModO dans ce cas) ne mettent pas

A

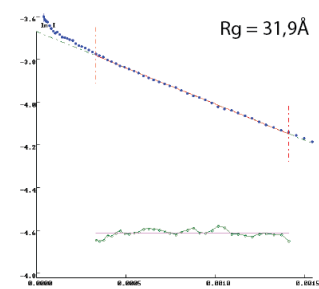
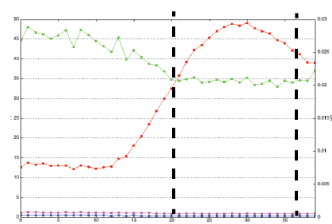
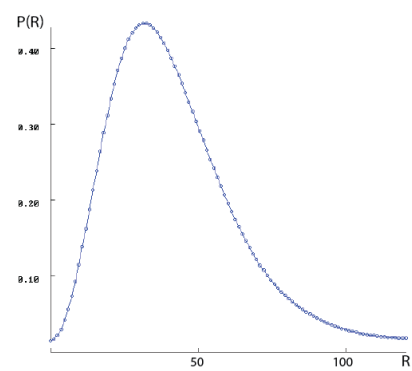
His-Rna15p

 $R_g = 38,5\text{\AA}$ **B**

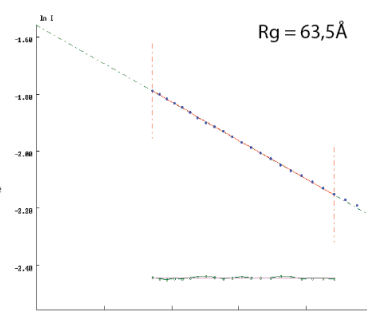
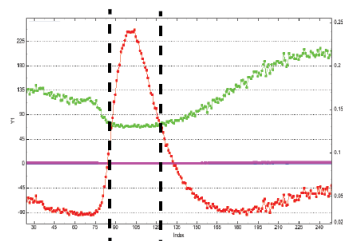
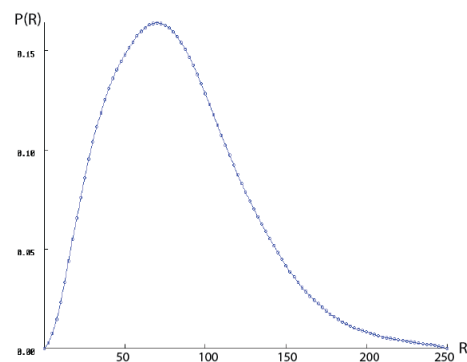
His-Rna15p Rna14p

 $R_g = 54,8\text{\AA}$ **C**

His-Pcf11p (454_563) Clp1p

 $R_g = 31,9\text{\AA}$ **D**

CF IA His-Clp1p Pcf11p (298-626)

 $R_g = 63,5\text{\AA}$ **E**

CF IA His-Rna15p Pcf11p (298-626)

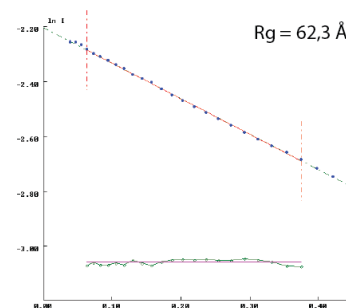
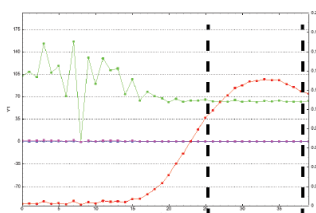
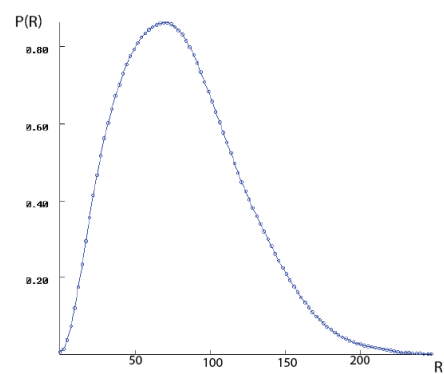
 $R_g = 62,3\text{\AA}$ 

Figure 30 : Illustration du processus d'acquisition et de traitement des données par le logiciel FOXTROT (développé localement). Le premier graphique représente le profil d'élution (rouge) à la sortie de la colonne HPLC BioSEC3 300 Å Agilent avec une estimation du rayon de giration (R_g , vert) associé à chaque prise d'absorbance à 280nm. Les deux tracés pointillés verticaux noirs indiquent les images sélectionnées pour poursuivre l'analyse. Le second graphique représente l'analyse de Guinier appliquée aux clichés sélectionnés. La courbe bleue représente l'intensité de la diffusion en fonction de l'angle de diffusion. L'adaptation à l'approximation de Guinier est représentée en rouge, la courbe verte représentant la différence entre la mesure et la prédiction. Une valeur plus précise du rayon de giration est calculée à cette étape. Le dernier graphique représente le calcul de la $P(R)$ pour l'échantillon analysé, avec R exprimé en Å. Cette illustration ne représente pas une liste exhaustive des nombreuses constructions testées et acquisitions effectuées.

en évidence de tendance dans le positionnement des parties flexibles. Il est important de comprendre que la diffusion théorique de toutes ces propositions corrèle très bien avec les données expérimentales. Et ces structures ne donnent pas les conformations qui existent en solution, seulement des modèles compatibles avec les données SAXS. Et compte-tenu de la grande variabilité de structures possibles, il apparaît que cette région est peu structurée en solution, donc invisible en cristallographie. Il est cependant également possible qu'une structure unique soit présente en solution, mais que son profil de diffusion ne soit pas suffisamment discriminant.

Nous avons également essayé de générer une enveloppe à partir de nos données. Une des modélisations est représentée en gris sur la Figure 31. Cependant, dans la mesure où l'on a une structure flexible et que le paramétrage du calcul d'enveloppe force cette dernière à être la plus compacte possible, nous avons une inadéquation *de facto* entre les calculs basse résolution et les calculs tout atome. Toute superposition entre les deux modélisations est donc inappropriée, et les calculs d'enveloppe sont ici peu informatifs.

La modélisation avec la meilleure corrélation est représentée en vert et bleu dans le panneau A, et les courbes sont représentées dans le panneau B. L'axe des X est une fonction de l'angle de diffusion et l'axe des Y est une fonction de l'intensité de diffusion. Les courbes de diffusion théoriques et expérimentales, respectivement oranges et bleues, sont presque parfaitement superposées. Les écarts étant indiscernables sur cette représentation, ils sont représentés par la courbe jaune.

Finalement, l'information que l'on peut dégager de cette analyse de His-Pcf11p (454-563) Clp1p est que l'on est bien en présence du complexe en solution, et que le SAXS confirme le caractère flexible des parties C-terminale et N-terminale du fragment de Pcf11.

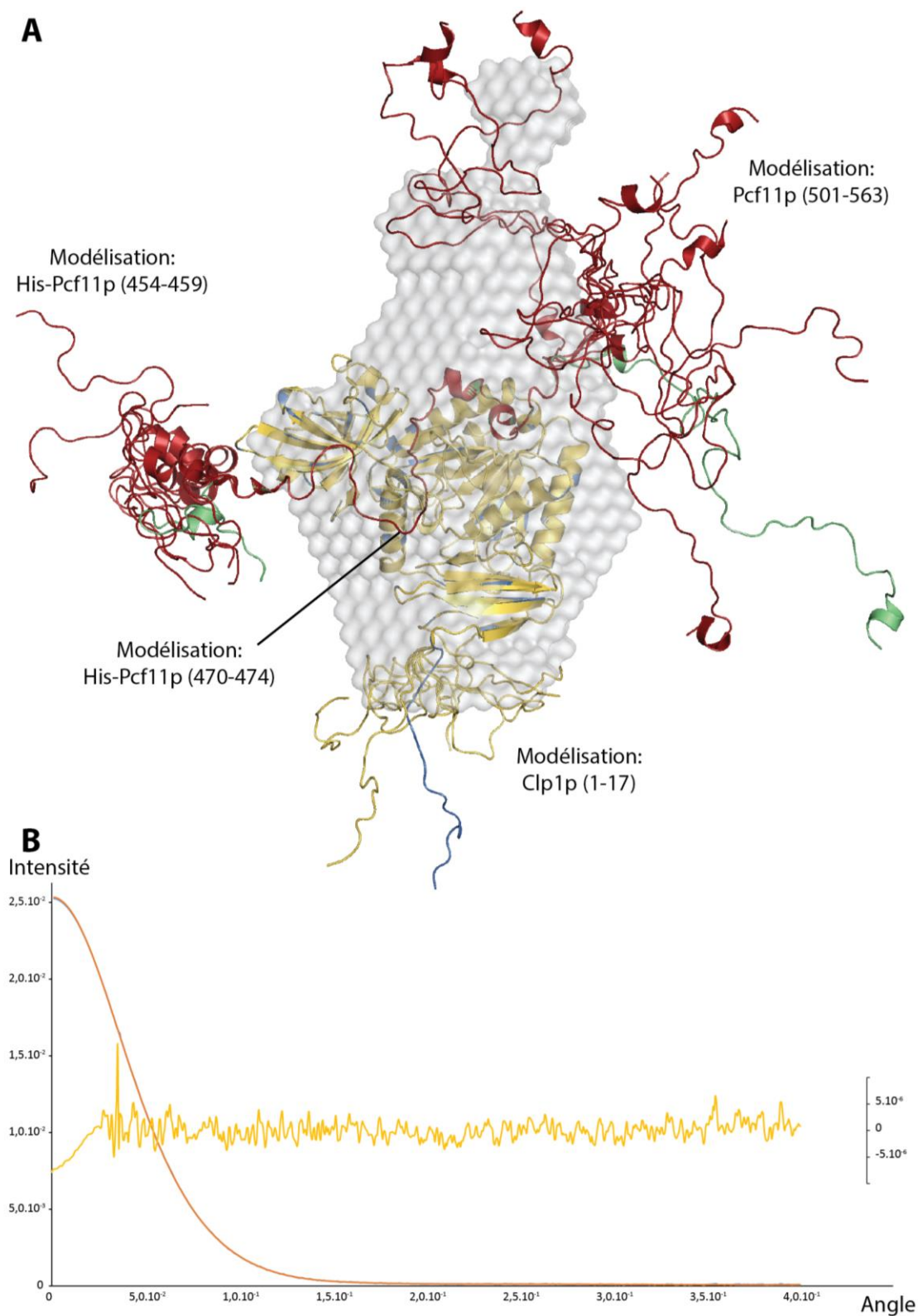


Figure 31 : Modélisations de la structure complète de His-Pcf11p (454-563) Clp1p. **A.** Représentation des différents essais de modélisation par DaDiModO de structures dont la diffusion théorique correspondrait à la diffusion mesurée. His-Pcf11p (454-563) est représenté en rouge (vert pour la meilleure solution) et Clp1p est représenté en jaune (bleu pour la meilleure solution). Un essai de modélisation d'enveloppe par DAMMIF est également représenté en gris. **B.** Comparaison de la diffusion théorique (orange) de la meilleure modélisation avec la diffusion mesurée (bleue). La différence absolue (jaune) entre les deux courbes est également représentée.

Nous avons eu la même démarche intellectuelle sur le complexe Rna15p-Rna14p. La Figure 32 présente les résultats de l'exploitation des données obtenues sur ces protéines. Deux modélisations sont représentées dans le panneau A, toutes deux en parfaite adéquation avec la diffusion expérimentale. La figure doit être lue en associant le dimère de Rna14p bleu aux deux monomères orange et jaune de Rna15p pour reconstituer la première modélisation. La seconde modélisation est constituée des fragments beiges (dimère de Rna14p), rouges et verts (monomères de Rna15p). Bien que leurs diffusions théoriques s'adaptent aux mesures, nous pouvons à nouveau remarquer que les deux modélisations sont considérablement différentes. Au vu de la très grande mobilité présente dans le complexe, nous n'avons pas jugé pertinent de faire des modélisations d'enveloppe (calculs basse résolution).

Le panneau B présente la diffusion théorique de la première modélisation (bleue – jaune – orange) en orange, par rapport à la diffusion expérimentale en bleue. Les écarts, rapportés sur la courbe jaune, semblent plus importants aux petits angles. Mais compte tenu des différences d'échelles, aucun écart mesuré entre 0 et 10^{-1} ne dépasse en fait la barre des 1% si on le rapporte à l'intensité mesurée.

Nous retrouvons finalement le même cas de figure qu'avec Pcf11p-Cpl1p. Nous possédons des données structurales pour une bonne partie de la séquence, mais les possibilités de modélisations restant en adéquation avec la diffusion expérimentale sont encore bien trop nombreuses.

Ma stratégie consistait initialement à caractériser les sous complexes du CF IA, plus simples, avant d'utiliser ces informations pour faciliter l'étude du facteur entier. Au vu de l'incapacité du SAXS à proposer, si ce n'est une solution claire, mais au moins une tendance dans l'organisation spatiale des protéines, nous avons renoncé à exploiter les données acquises sur le facteur. Quelques essais rapides (données non montrées) par acquis de conscience montrant une vaste diversité dans les modélisations basses résolution nous ont confortés dans cette décision.

3.2.4.4. Conclusions

Finalement, bien que nous ayons réussi à acquérir de bonnes données sur les échantillons homogènes, peu d'information concrète a pu en être déduit. En effet, les

différentes modélisations effectuées ne mettent pas en évidence de tendance dans le positionnement des parties flexibles. Les nombreux modèles générés font tous état d'une forte mobilité, mais sans que nous ayons la certitude qu'ils aient une quelconque réalité. Nous savons juste qu'ils sont compatibles avec les mesures effectuées. Ceci est une limite intrinsèque à la technique, à laquelle on ne peut se soustraire, mais la mise en évidence d'une tendance aurait amené des informations.

La liberté actuellement conférée à la modélisation est trop importante, mais les données sont acquises. Nous pourrions y revenir à tout moment si nous arrivons, en cristallographie, microscopie électronique ou tout simplement par tests d'interaction, à mettre en évidence de nouvelles contraintes qui contribueraient à limiter les possibilités de la modélisation.

3.2.5. Microscopie électronique

Le travail important fourni dans le cadre de ces travaux de thèse autour de la technologie de la microscopie électronique a été effectué en collaboration avec Patrick Bron et Joséphine Lai Kee Him, de l'équipe « Biologie Structurale Multi-Echelles » du Centre de Biochimie Structurale de Montpellier.

Visant les mêmes objectifs que l'approche de SAXS, c'est-à-dire la détermination d'une enveloppe à basse résolution, nous ne sommes cependant pas ici en terrain complètement inconnu. En effet, Noble et al. ont publiés en 2004 un article (Noble, Walker et al. 2004) explorant les rôles de Rna14p et Rna15p dans la liaison à l'ARN. Cet article est le premier à reconstituer l'hétérocomplexe Rna14p-Rna15p à partir de protéines recombinantes, et ils utilisent notamment ce matériel pour une étude par microscopie électronique à transmission, en coloration négative (Figure 33). Des particules sont observées, allongées, avec un angle de 110° dessinant un bras long (100\AA) et un bras court (80\AA). Cependant, ces observations ne leur permettent pas de proposer un modèle plus poussé mais seulement des croquis de ces objets qu'ils attribuent à un hétérodimère de Rna14p-Rna15p grâce à un calcul du volume moyen de leurs objets (objets : 125000\AA^3 ; hétérodimère théorique : 138000\AA^3). Nous pouvons cependant remarquer qu'ils identifient ces objets à hauteur de 60% des particules identifiées. Le reste semble être constitué par des associations, lié par la pointe du bras court selon des angles variables, correspondant donc à des hétérooctamères (2 : 2 + 2 : 2).

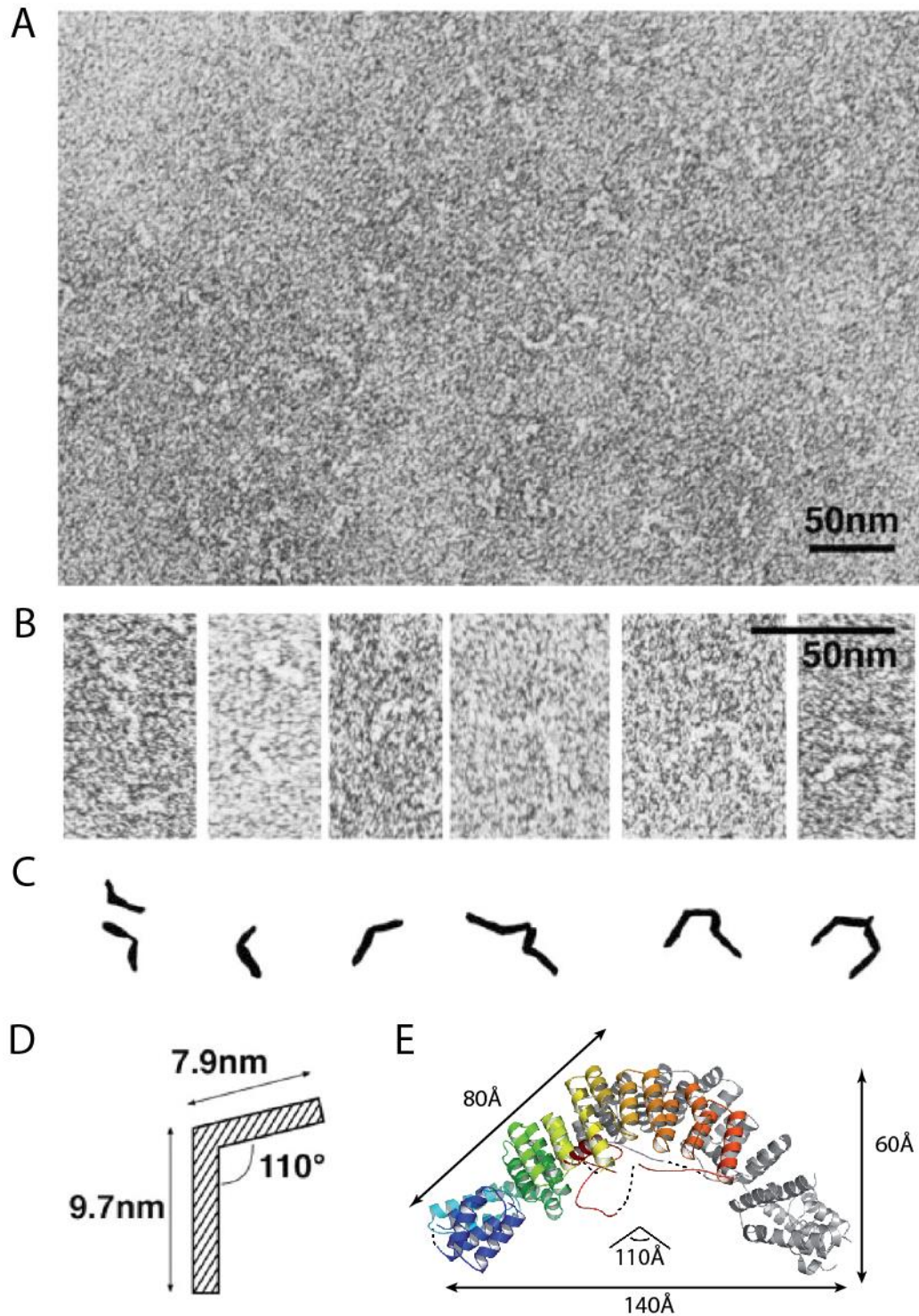


Figure 33 : Données de microscopie électronique obtenues par Noble et al. en 2004. A. Clichés de coloration négative sur les échantillons de Rna14pRna15p recombinants. B. et C. Isolement et schématisation des objets. D. Modèle proposé E. Structure et dimensions du dimère des HAT-repeat de Cstf-77ec (Legrand, Pinaud et al. 2007) (PDB : 2UY1)

Les structures de CstF-77 (homologue de Rna14p) publiées en 2007 par le laboratoire (Legrand, Pinaud et al. 2007) et Bai et al. (Bai, Auperin et al. 2007) viendront apporter des éléments corrélant avec ces observations. En effet, les structures ne contiennent que

l'homodimère des domaines HAT, mais un angle de 110° est formé par le dimère (Figure 33). Ils observent également que chaque bras de leur structure fait environ 80Å, longueur retrouvée pour le bras court des images de microscopie électronique. En effet, bien que Noble et al. (Noble, Walker et al. 2004) proposent que ces objets soient des hétérodimères, ils ne discutent pas du tout la présence d'un bras long et d'un bras court, là où nous aurions pu attendre une symétrie.

Depuis, aucune autre observation de microscopie électronique de CF IA ou de l'une de ces protéines n'a été publiée, laissant inexploré le champ de l'observation directe de CF IA.

3.2.5.1. Principe

La microscopie électronique est un champ de la microscopie utilisant un faisceau d'électron pour visualiser l'échantillon. Comme la longueur d'onde des électrons est 100 000 fois plus petite que celle des photons de la lumière visible, les résolutions et les magnifications possibles sont bien plus importantes qu'avec un microscope photonique. Il existe alors deux stratégies pour obtenir une image de notre échantillon : la transmission, ou le balayage. Dans les microscopes à transmission, comme celui que nous avons utilisé durant ce travail, le faisceau d'électrons traverse l'échantillon positionné sur une grille de carbone, et l'image est créée par la capacité des atomes de l'échantillon à bloquer le faisceau. Concrètement, plus le nuage électronique rencontré par le faisceau est important, plus les électrons de la source auront de probabilité d'être déviés, donnant un pixel noir. A l'inverse, une matière peu dense permettra le libre passage rectiligne des électrons qui activeront le détecteur et donneront un pixel blanc.

Une image obtenue en microscopie électronique par transmission correspondra donc à une projection en deux dimensions de notre objet, en fonction de son orientation sur la grille de cuivre. Un même objet donnera alors une multitude d'images, plus ou moins similaires selon sa symétrie. Lors de l'acquisition des images, nous chercherons donc à observer des formes distinctes qui reviendraient régulièrement. Puis, lors du traitement, le logiciel cherchera à extraire des différents clichés des formes spécifiques se détachant du bruit afin de les classer et regrouper celles qui correspondent à une projection de l'objet dans la même orientation. En moyennant ainsi les images similaires, on améliore le contraste et on peut à

ce moment essayer de reconstituer un modèle en trois dimensions cohérent avec les projections 2D obtenues.

3.2.5.1. Echantillons testés

L'idée globale est d'obtenir une enveloppe à basse résolution d'un complexe CF IA le plus complet possible. Une enveloppe permettant de situer les structures connues serait déjà extrêmement informative, et des constructions délétées de certaines parties flexibles sont peut-être moins susceptibles d'être hétérogènes et donnent des alternatives intéressantes pour localiser les domaines.

C'est dans cette optique que les complexes purifiés et expédiés ont été sélectionnés. Voici la liste exhaustive des échantillons analysés :

- HisRna15p-Rna14p
- CF IA *S. cerevisiae*
 - His-Clp1p + Pcf11p (298-626) + Rna14p-Rna15p
 - His-Clp1p, Pcf11p (331-563) + Rna14p-Rna15p
 - His-Pcf11p (298-626) + Rna14p-Rna15p + Clp1p
 - His-Pcf11p (331-626) + Rna14p-Rna15p + Clp1p
 - His-Pcf11p (331-563) + Rna14p-Rna15p + Clp1p
 - His-Pcf11p Δ Q + Rna14p-Rna15p + Clp1p
 - His-Pcf11p Δ Q + Rna14p-Rna15p + Clp1p (GraFix)
- CF IA *K. lactis*
 - His-Pcf11kl (322-518) + Rna14kl-Rna15kl + Clp1kl

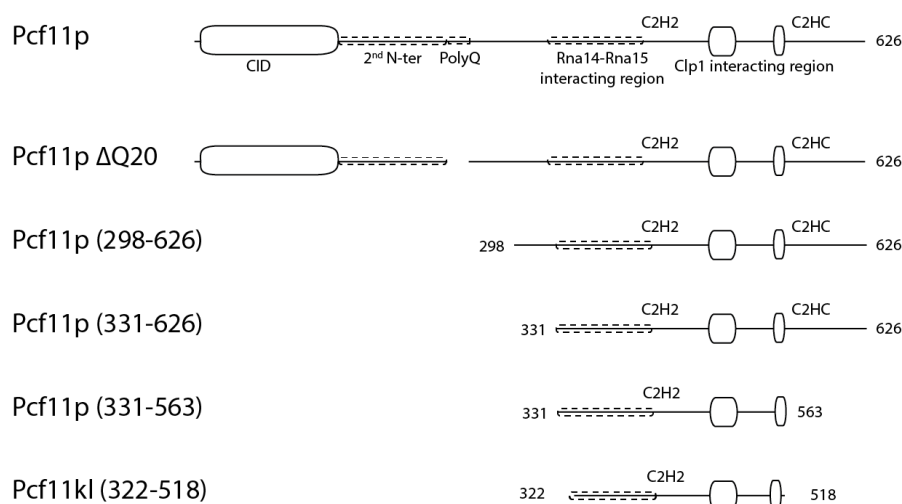


Figure 34 : Représentation schématique des constructions de Pcf11 utilisées pour l'expression des complexes CF IA analysés en microscopie électronique. Pcf11p est représenté pour servir de référence.

La figure ci-dessus répertorie les différentes constructions de Pcf11 explorées. En termes de séquences, on voit bien que nous avons essayé de nous concentrer autour d'un Pcf11 « minimal », compétant pour les interactions reconstituant CF IA. J'ai également essayé plusieurs conditions de préparation des échantillons, notamment en jouant sur la concentration et la congélation, en congelant avec ou sans glycérol. Suite à plusieurs problèmes avec le prestataire responsable du convoyage des échantillons, nous avons également mis sur pied un déplacement d'une semaine pendant lequel je me suis rendu à Montpellier avec des échantillons qui n'ont alors pas été congelés.

3.2.5.1.1. Coloration négative

Le principe de la coloration négative est contre intuitif : il s'agit de colorer l'environnement, et non l'échantillon. En effet, la capacité à être opaque aux électrons est directement reliée au numéro atomique des atomes, c'est-à-dire à leur nombre de protons. Or la majorité de la matière biologique est constituée d'atomes légers, avec un faible pouvoir diffuseur. La coloration négative consiste donc à imprégner l'échantillon avec un colorant qui présente deux caractéristiques : diffuser fortement les électrons et être adsorbé par les échantillons biologiques. Les endroits adsorbant le colorant augmenteront d'autant leur opacité aux électrons, facilitant l'observation et améliorant le contraste. Les colorants seront

donc préférentiellement composés d'éléments lourds. Durant les différentes acquisitions, nous avons testés trois adsorbats différents :

- Molybdate d'ammonium
- Acide phosphotungstique
- Acétate d'uranyle

L'objectif de la coloration négative est de créer autour des échantillons une couche homogène d'une substance dense aux électrons afin que les structures apparaissent en négatif : clair sur fond sombre.

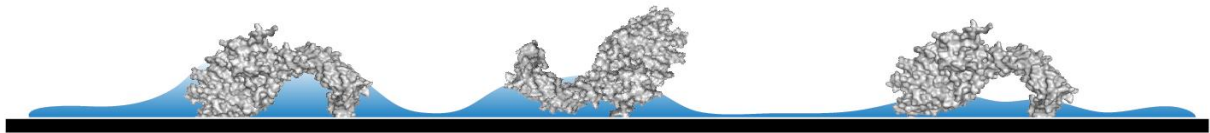
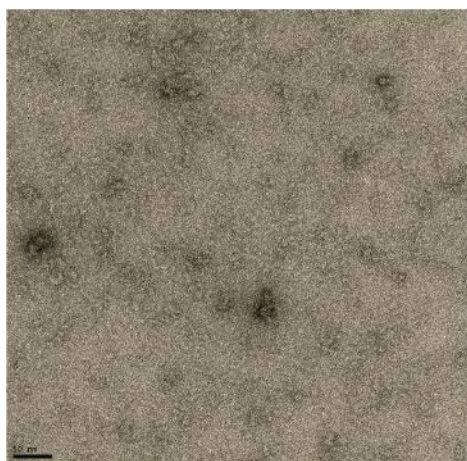


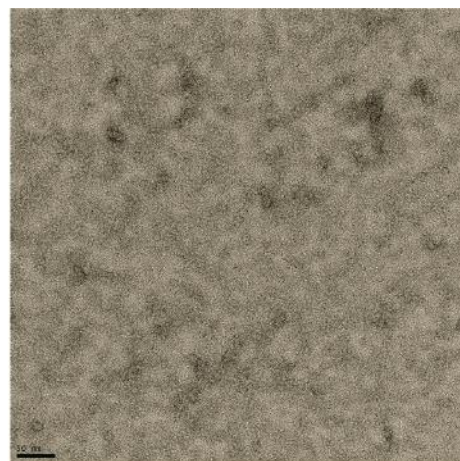
Figure 35 : *Principe de la coloration négative. Le colorant (représenté en bleu), forme une couche uniforme, sauf à proximité des échantillons, augmentant ainsi le contraste.*

Durant le projet, nous avons testé tous les échantillons en coloration négative, car cela est rapide à expérimenter, et Patrick Bron considère, par expérience, qu'un échantillon difficile à exploiter en coloration négative ne donnera rien en microscopie cryo-électronique.

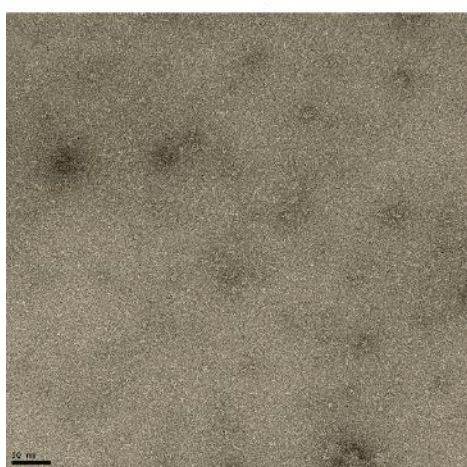
La sélection de clichés présentés dans la figure ci-dessus n'est pas exhaustive des conditions testées avec le CF IA ΔQ mais elle reste représentative. Nous n'avons en effet pas été capables d'identifier de condition de dilution et de colorant permettant de discerner des objets isolés. A la place, nous observons une couche de matériel indistinct. Des dilutions supplémentaires, même très légères, conduisent directement à une grille sur laquelle on n'observe plus de protéine.



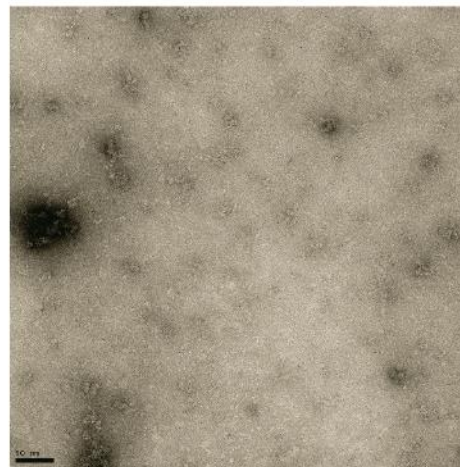
CF IA ΔQ 0,1 mg/ml
Acétate d'uranyle 1%



CF IA ΔQ 0,07 mg/ml
Acétate d'uranyle 1%



CF IA ΔQ 0,025 mg/ml
Ammonium Molybdate



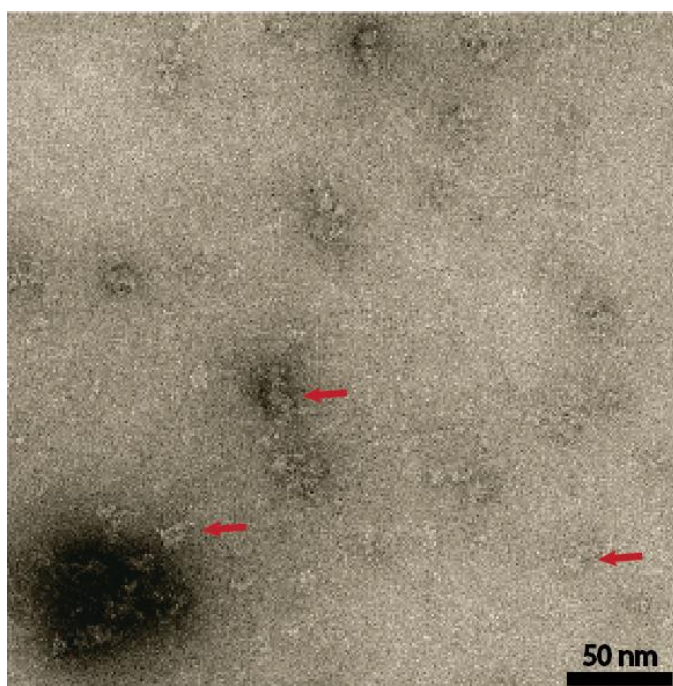
CF IA ΔQ 0,025 mg/ml
Acide phosphotungstique

Figure 37 : Exemples de clichés obtenus avec un échantillon de CF IA délété de la séquence de glutamine. Nous avons testé différentes concentration en facteurs et différents agent contrastant. Images acquises par Patrick Bron (CBS, Montpellier)

Ces remarques doivent toutefois être nuancées par l'image acquise à 0,025 mg/ml et colorée à l'acide phosphotungstique. Dans ce

Figure 36 : Agrandissement du cliché « CF IA ΔQ 0,025 mg/ml ; Acide phosphotungstique) de la figure précédente. On discerne ici des marques plus claires se détachant légèrement du bruit de fond, possiblement des complexes CF IA (flèches rouges)

cliché, un zoom nous permet



effectivement de remarquer la présence d'objets entre 100-150Å. Cependant, nous n'avons pas été capables d'améliorer le contraste ou la visibilité de ces objets et ils semblent ici fortement hétérogènes. Un traitement informatique de ces clichés n'a pas su identifier des objets exploitables.

L'observation d'objets, bien que ceux-ci soient indistincts, nous encourage donc à aller dans le sens de notre stratégie visant à réduire les zones flexibles de Pcf11. Nous avons donc testé d'autres constructions, en essayant à chaque fois différentes concentrations et différents agents contrastants.

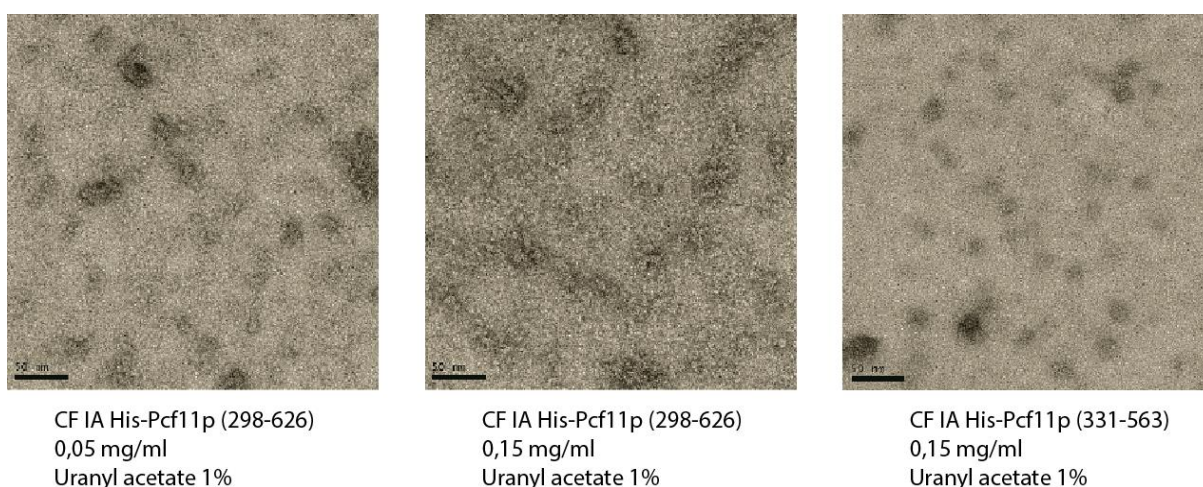


Figure 38 : Clichés illustrant de manière non exhaustive les essais de coloration négative effectués avec les différentes constructions de CF IA. Images acquises par Patrick Bron (CBS, Montpellier)

Nous pouvons voir dans la sélection de clichés présentés dans la figure précédente que nous ne sommes pas parvenus à obtenir d'image claire permettant de distinguer des objets. En effet, quelle que soit la construction, le contrastant ou la dilution essayée, nous ne sommes pas arrivés à observer autre chose qu'un tapis de protéine ou une grille vide. Nous avons obtenus les mêmes observations en nous intéressant à CF IA *K. lactis*.

3.2.5.1.2. Microscopie cryo-électronique

Le problème majeur de la coloration négative est qu'elle implique beaucoup de manipulations de l'échantillon ainsi que l'exposition à des substances chimiques pouvant potentiellement le dénaturer. Pour pallier à ces inconvénients, comme nous savons grâce à nos expériences de SAXS que nos échantillons sont homogènes et monodisperses, nous avons également essayé la microscopie cryo-électronique.

Cette stratégie consiste à intégrer l'échantillon dans de la glace vitreuse par congélation rapide. Ainsi, l'échantillon resté dans son tampon ne subit pas de modification de son environnement et on évite notamment la déshydratation ou l'altération de la structure. Cependant, cette approche étant dépourvue d'étape de coloration, le contraste est moins bon et le bruit beaucoup plus important.

Compte tenu du comportement des échantillons de CF IA en coloration négative, nous n'avons pas réussi à identifier de condition de concentration en protéine qui serait susceptible d'offrir de bonnes conditions d'observation. Nous avons donc fait quelques essais de microscopie cryo-électronique, qui se sont révélés infructueux. Un autre moyen pour explorer CF IA par microscopie électronique consiste à s'intéresser à Rna14 et Rna15. Dans l'optique d'obtenir plus d'information sur cet hétérodimère, nous avons donc préparé plusieurs purifications d'échantillon de complexe Rna14p-Rna15p, afin de tester la concentration et la méthode de congélation (avec ou sans glycérol). Cependant, bien que nous ayons essayé de reproduire et améliorer les images obtenues par Noble et al. (Noble, Walker et al. 2004) par coloration négative, nous ne parvenons pas à mettre en évidence d'objets sur nos clichés de cryo-microscopie (cf figure ci dessous).

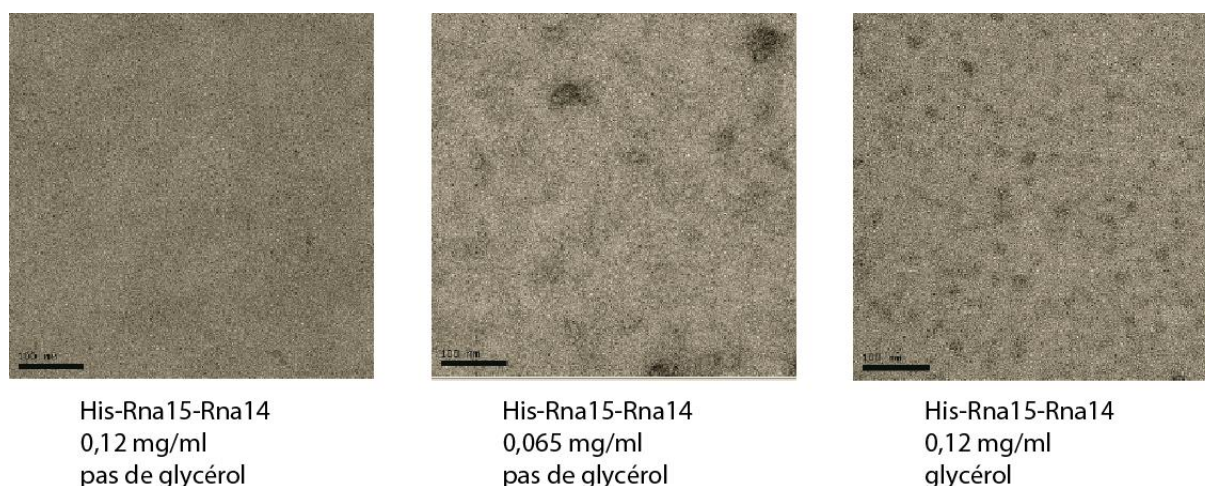


Figure 39 : Sélection de clichés obtenus avec les échantillons de Rna14p-Rna15p. Images acquises par Patrick Bron (CBS, Montpellier)

3.2.5.1.3. GraFix

Le GraFix (Gradient Fixateur) est une approche pour explorer et limiter la mobilité au sein du complexe. Nous avons en effet vu en SAXS que si les échantillons étaient homogènes, il est possible qu'ils présentent une mobilité intra moléculaire forte. Cette idée est cohérente avec la présence de nombreuses régions flexibles au sein de CF IA. Nous avons donc cherché une stratégie permettant d'explorer cette question, et le GraFix est une technique qui, couplée à la microscopie électronique, semble pouvoir convenir.

Le GraFix est décrit en 2008 par Kastner et al. (Kastner, Fischer et al. 2008). Cette technique consiste à faire migrer l'échantillon à travers un gradient de glycérol et d'agent pontant, dans notre cas du glutaraldéhyde. La migration est effectuée par centrifugation, afin que les complexes rencontrent une concentration croissante d'agent fixateur qui stabilisera les macromolécules individuelles. La migration au sein du gradient de glycérol permettra ainsi de séparer les objets selon leur taille mais également de limiter les interactions inter-objets. Le GraFix présente cependant deux principales limites vis-à-vis de la microscopie électronique. La première est intrinsèque : à l'issue du GraFix, l'échantillon est dans un milieu riche en glycérol, ce qui n'est pas adapté à des analyses de microscopie cryoélectronique, et nous devons donc ajouter des étapes de nettoyage de l'échantillon. La seconde est matérielle : pour limiter les interactions inter-facteurs au sein du gradient, la quantité et le volume de matériel que nous pouvons charger sur le gradient est limitée, et le passage dans le gradient va également diluer l'échantillon. Nous serons donc souvent obligés de reconcentrer le matériel pour pouvoir travailler.

La mise en place de la technique de GraFix pour une application à CF IA nécessitera cependant une mise au point relativement fastidieuse afin de déterminer les conditions de centrifugations optimales pour obtenir CF IA au milieu du gradient. Cette mise au point fut effectuée par Lionel Minvielle-Sébastien, qui souhaite à terme effectuer des essais avec du CF IA endogène. Il me fallut uniquement adapter mes conditions de tampon pour la dernière étape de chromatographie d'exclusion afin de m'affranchir de la présence du Tris et de passer en tampon HEPES. Cela me permit de procéder rapidement à des essais avec le CF IA recombinant, une seule purification me permettant d'obtenir suffisamment de matériel pour charger plusieurs gradients.

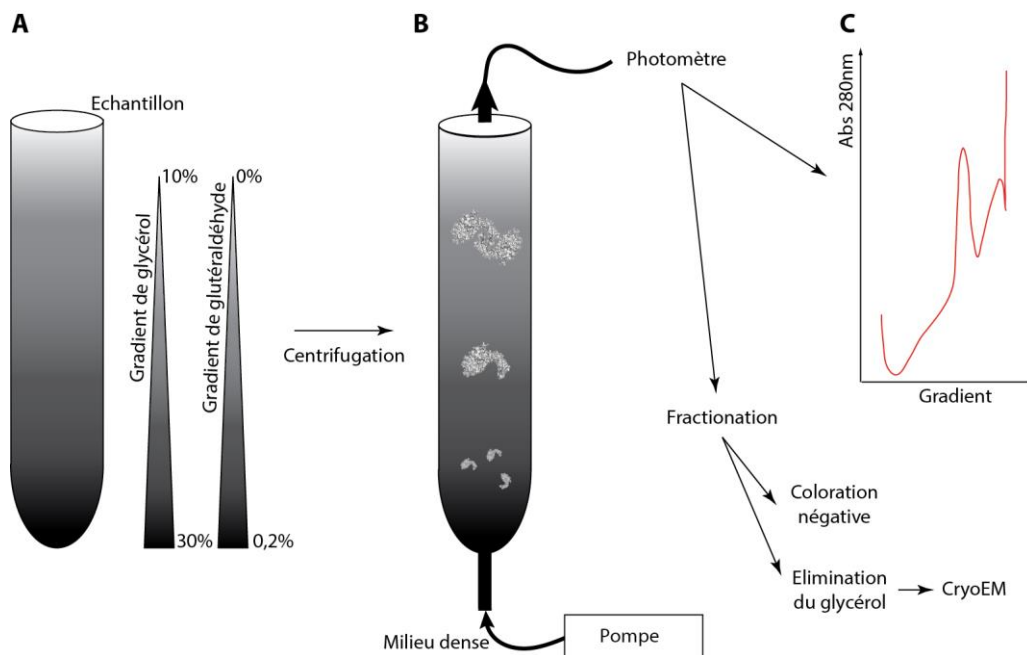


Figure 40 : Description de la technique de GraFix (adapté de Kastner, 2008). A. La première étape consiste à générer des gradients de glycérol et de glutéraldéhyde, à la surface desquels l'échantillon va être déposé. B. Une étape de centrifugation va permettre à l'échantillon de pénétrer dans les gradients et s'y répartir selon les vitesses de sédimentation de ses composants. C. Le matériel sera alors récupéré sans perturber le gradient en injectant un milieu dense.

Le dosage au nanodrop des deux fractions de la tête de pic de gel filtration m'indique que la fraction 25 est à 0,11 μM tandis que la fraction 26 est à 0,33 μM , et nous voyons sur le gel (Figure 41, A) qu'elles sont parfaitement propres. J'ai donc décidé de les charger tel quel sur le gradient, à hauteur de 200 μL d'échantillon par tube de gradient, ce qui représente respectivement 22 et 66 pmol pour chaque échantillon. Trois gradients seront fait avec chaque échantillon. L'un sera systématiquement sans glutaraldéhyde (graphique B, à gauche) et nous servira à faire différents contrôles. Dans un premier temps, il sert à régler les paramètres d'acquisition du photomètre, comme en témoignent les sauts dans les courbes causés par les réglages. Après une précipitation TCA des fractions, nous pourrons les charger sur un gel SDS-PAGE (illustré en C), qui nous permettra de calculer le délai de sortie entre le pic enregistré par le photomètre et la collection des fractions, qui est faite à la main. En identifiant ce délai, qui s'est avéré être d'une fraction (environ 100 μl), nous sommes alors en mesure d'indiquer avec précision à Patrick Bron dans quelles fractions se situe le facteur. Un dosage effectué sur ces précipitations nous permettra également de déterminer la concentration du facteur dans les fractions du pic, information nécessaire pour le travail de préparation des grilles de microscopie cryo-électronique.

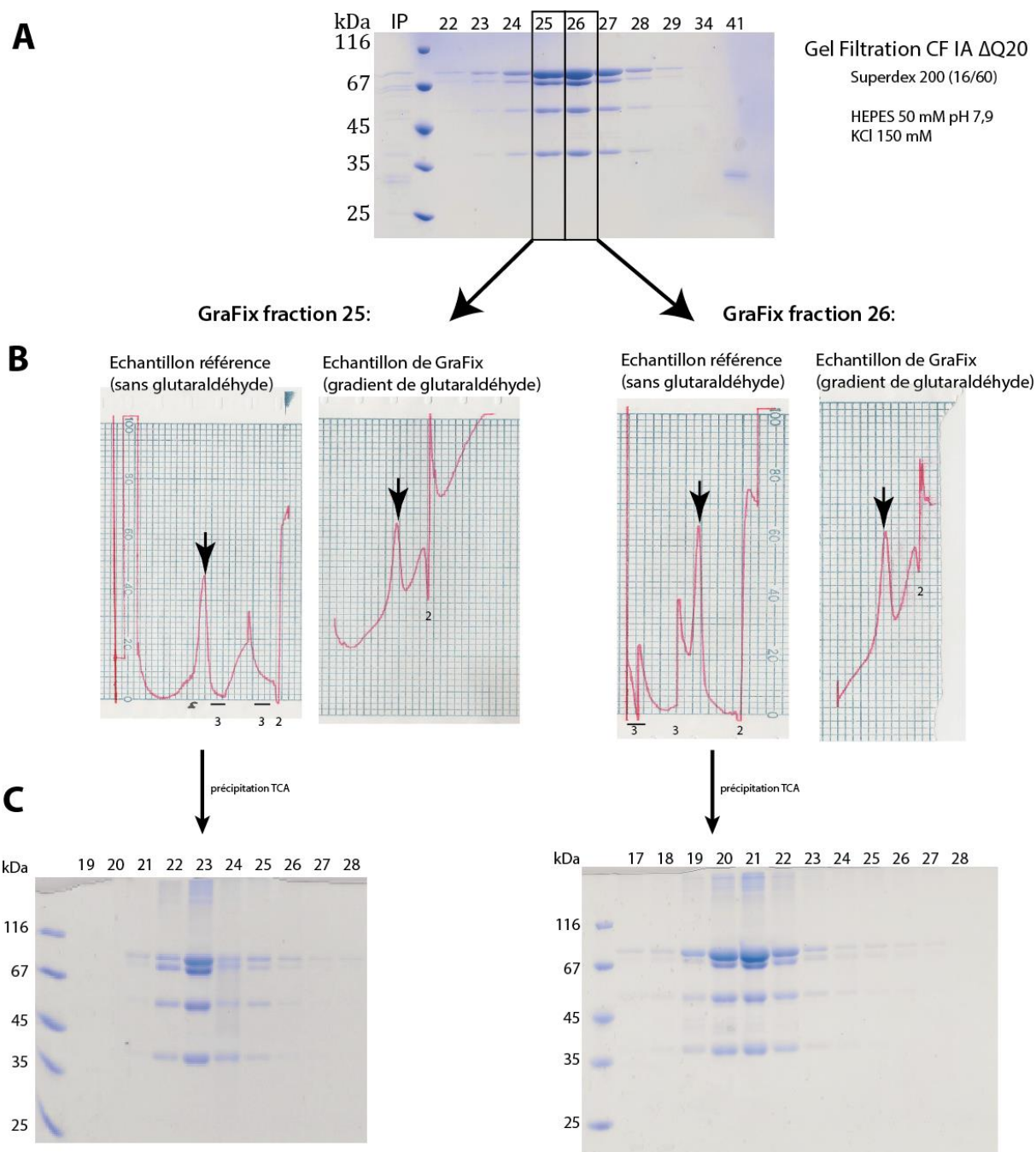


Figure 41 : GraFix avec du CF IA Δ Q. A. Gel SDS-PAGE (12%) montrant l'échantillon en sortie de purification. B. Chromatogrammes obtenus à l'issue des centrifugations. Pour chaque échantillon le contrôle natif est à gauche et le GraFix est à droite. La flèche souligne le pic d'absorbance créé par l'échantillon, 2 indique la fin du gradient et 3 indique des décrochages de la courbe dus à des réglages. C. Gel SDS-PAGE (12%) sur les échantillons natifs après précipitation TCA afin de vérifier la qualité de l'échantillon et de déterminer le délai de sortie.

Si on s'intéresse au profil de sédimentation obtenu avec les échantillons de CF IA Δ Q (B), on observe que les pics, uniques, (B, flèches noires) se dessinent de manière très claire au milieu de la progression régulière formée par le gradient. Le décrochage final rencontré par la

courbe et signalé par un « 2 » correspond à la fin du gradient et au passage devant le détecteur du ménisque formé par le milieu dense (60% glycérol) que nous utilisons pour pousser le gradient hors du tube. En absence ou en présence de glutaraldéhyde, aucun pic secondaire n'apparaît, indiquant qu'aucune espèce de masse inférieure ou supérieure à celle de CF IA n'est stabilisée dans le fixateur. Le pic de CF IA lui-même montre un profil gaussien, que l'on retrouve un peu sur les gels (C), et qui indique un échantillon bien homogène.

Les fractions réticulées sont alors simplement congelée en azote liquide et expédiées à Patrick Bron. Quelques mises au point furent nécessaires pour éliminer correctement le glycérol, mais finalement, aucune condition permettant d'observer des objets nets ne put être identifiée.

Toujours avec la volonté d'améliorer les schémas proposés par Noble et al. (Noble, Walker et al. 2004), Lionel Minvielle-Sébastien a également préparé sur Grafix, des échantillons de Rna15p-Rna14p recombinant que je lui ai fournis et qui ont été envoyés à Patrick Bron. De belles acquisitions ont pu être faites sur ces échantillons (Figure 42, A) dans lesquelles on distingue très nettement des objets. Après traitement informatique et isolation des objets (B), il s'avère que nous avons à faire à des profils plus ou moins ronds, d'environ 120 Å de diamètre, ou légèrement allongés, avec jusqu'à 150 Å de long pour 120 Å de large (C). Trop hétérogènes, Patrick Bron n'est pas parvenu à regrouper les objets identifiés en classes pour en améliorer la résolution et essayer de construire un modèle.

Cependant, nous pouvons comparer les dimensions estimées de nos objets avec une structure publiée quelques mois après le début de notre projet par Paulson et Tong (Paulson and Tong 2012). Dans cette publication, ils présentent la structure cristalline du complexe Rna14 K lactis - Rna15 K lactis à 3.3 Å. Obtenue par la cristallogénèse d'échantillons de Rna14kl-Rna15kl (idem) recombinants entiers, ils n'observent finalement que les HAT-repeat de Rna14kl et l'interaction hinge - monkey tail décrite en 2011 (Moreno-Morcillo, Minvielle-Sébastien et al. 2011). Par contre, dans leur structure, et contrairement aux structures déjà résolues, ils observent une organisation hexamérique des protéines (Figure 42, D). Cette organisation dessine un cylindre d'environ 120 Å de diamètre pour 150 Å de long, ce qui est bien plus volumineux que les objets observés par Noble, mais qui pourrait correspondre aux objets que nous observons (C).

La question qui se pose alors est : pourquoi observons-nous des objets qui semblent être hexamériques en microscopie quand le comportement de notre échantillon durant la purification (gel filtration ou même SEC-MALS) est plus caractéristique d'hétérodimères? Noble et al. (Noble, Walker et al. 2004), en soutien à leurs observations de microscopie, font également des mesures de DLS et d'ultracentrifugation analytique, venant appuyer le fait que leur échantillon ne soit en aucun cas composée d'espèce aussi massive qu'un hétérohexamère. Pour clarifier la stœchiométrie de nos échantillons, nous avons décidé de faire des analyses de spectrométrie de masse en conditions natives (cf chapitre 3.2.1.2.1). Et nos résultats montrent clairement une espèce correspondant à une stoechiométrie 2 :2, et donc la formation d'un hétérodimère. Aucune espèce de masse moléculaire supérieure n'est détectée dans les conditions de l'analyse, nous laissant penser que si hexamère il y a, il est soit minoritaire, soit relativement peu stable. Dans tous les cas, si nous pouvons expliquer la présence d'un hexamère dans la structure par la pression de l'empilement cristallin, nous sommes incapables d'expliquer l'observation d'hexamères sur nos clichés, s'il s'agit effectivement bien d'hexamères.

Cependant, indépendamment de la nature des objets observés, on constate que malgré un pontage avec le GraFix, nous conservons une forte hétérogénéité structurale, trop importante pour permettre l'analyse. Avec le GraFix, nous espérons contrôler la mobilité des régions flexibles et stabiliser le complexe dans ses agencements les plus communs, tout en augmentant sa stabilité pour faciliter et les manipulations et le traitement des données, mais cela n'a finalement pas été le cas.

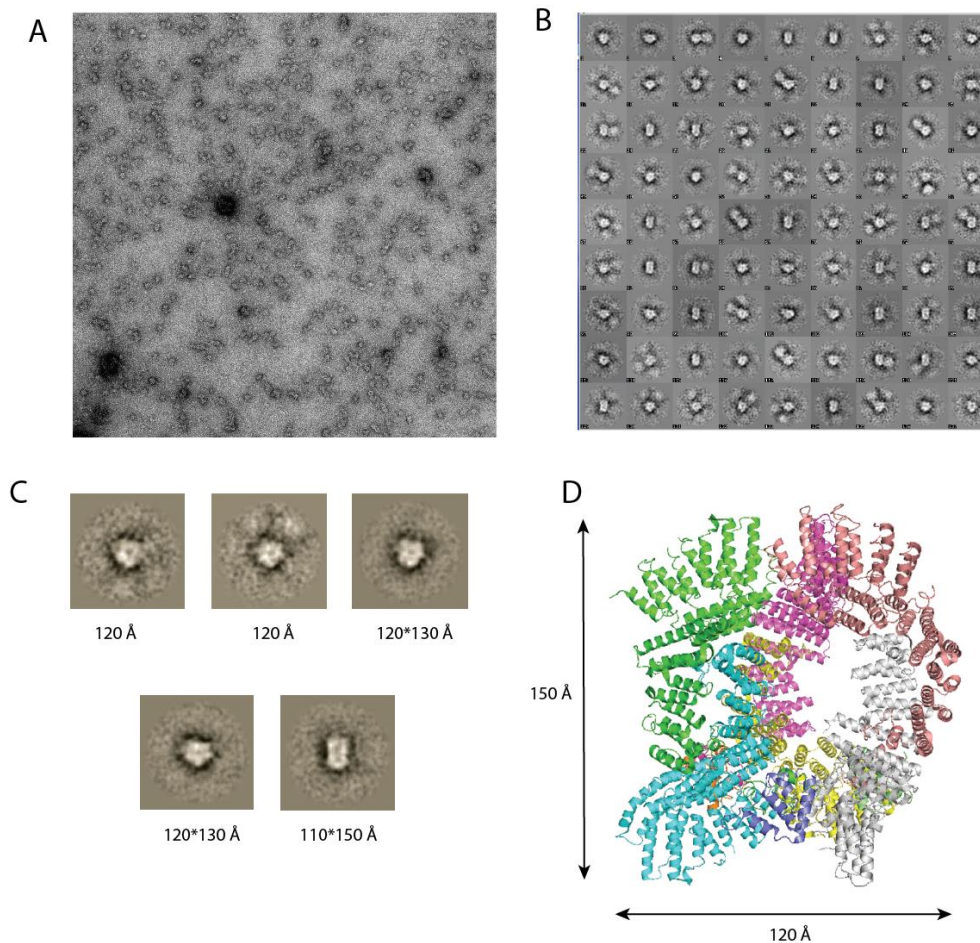


Figure 42 : Microscopie cryoélectronique sur le complexe *Rna14p-Rna15p* recombinant. A. Cliché de l'échantillon (Patrick Bron, CBS) B. Exemples d'objets extraits des séries de clichés. C. Regroupement des objets par classes et dimensions. D. Structure et dimensions de l'hexamère *Rna14kl-Rna15kl* résolu par Paulson et al. (Paulson and Tong 2012) (PDB : 4EBA ; *Rna14kl* sont en jaune, violet, vert, cyan, saumon et gris. *Rna15kl* sont en orange, vert pâle et bleu)

3.2.5.1.4. Cristallisation 2D

La cristallisation 2D est une technique un peu particulière. Elle consiste à essayer de forcer la formation de réseaux bidimensionnels par la molécule d'intérêt ancrée dans une bicouche lipidique (Figure 43). Ces cristaux seront donc dans un plan, l'ancrage de chaque molécule dans la bicouche empêchant le développement d'une troisième dimension. Mais cette organisation pourra être observée en microscopie, et la symétrie créée par l'organisation cristalline permettra de faciliter la localisation de l'objet ainsi que le traitement des images. On peut également imaginer faire de l'ensemencement (seeding) avec ces cristaux pour essayer de leur faire développer une troisième dimension et se tourner vers la diffraction des rayons X.

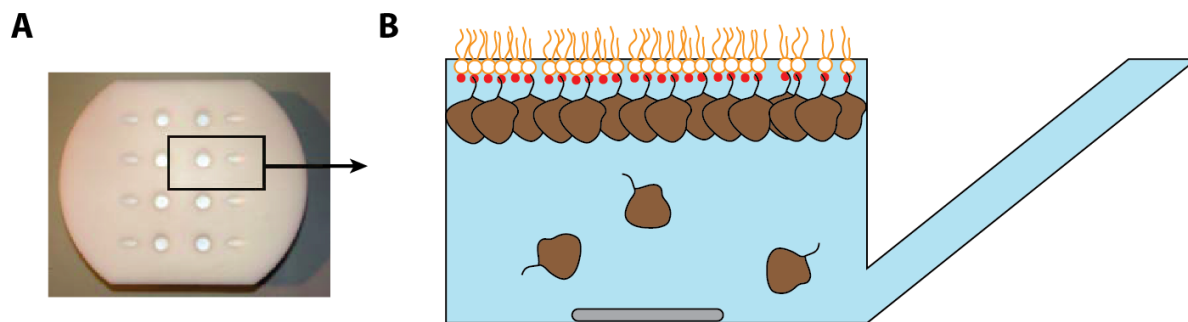


Figure 43 : Principe de la cristallisation 2D. A. La plaque en téflon possède 8 puits de travail. B. Représentation d'une cristallisation 2D. Une monocouche lipidique se forme à l'interface. Les ions (en rouge) fixés aux têtes lipidiques permettent l'ancrage des protéines His-tagguées (en marron). Le canal latéral sert à injecter les protéines sans perturber les lipides.

Cette technique présente l'avantage de conserver la protéine dans son tampon tout le long de l'expérience. Nous ne rajouterons que des ions métalliques afin de tenter de créer des conditions favorables à une organisation cristalline. Le réservoir est maintenu sous agitation douce. La monocouche lipidique est formée sur la nuit, puis l'échantillon est ajouté par le canal latéral. Nous laisserons quelques heures d'incubation pour permettre la formation et l'organisation des complexes lipides-protéines, puis une grille de microscopie électronique sera déposée sur les lipides pour récupérer la monocouche qui sera observée en coloration négative.

J'ai pu essayer de nombreux complexes différents, avec différents sels durant ma mission au CBS, mais sur aucune des images analysées je n'ai pu remarquer de réseau cristallin. Comme on peut le voir sur les clichés présentés en Figure 44, nous avons une couche protéique de formée (CF IA – His-Pcf11p (298-626) en CaCl_2) mais sans organisation. Dans d'autres cas, nous pouvons observer des agrégations très marquées.

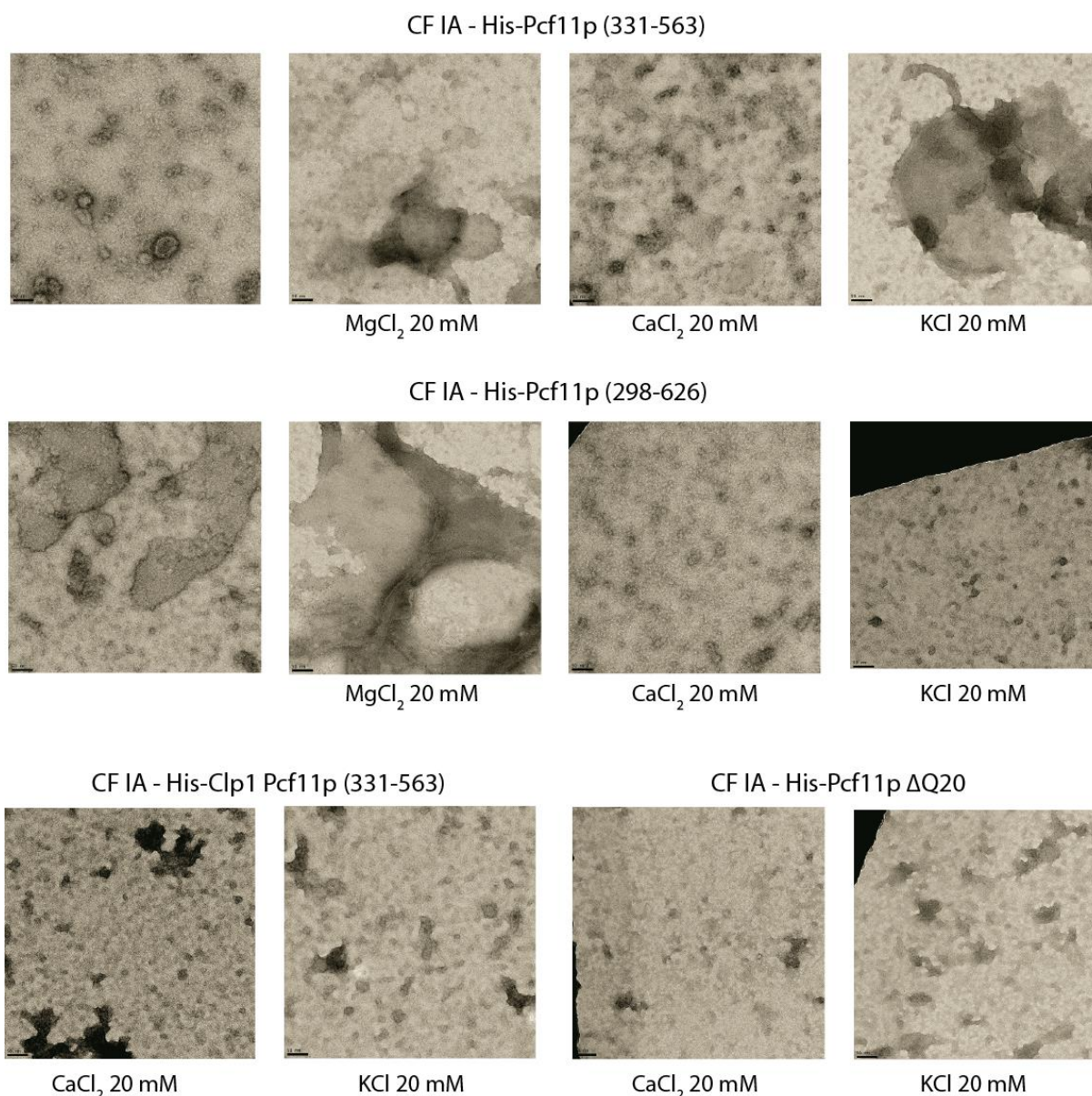


Figure 44 : Exemples d'observation d'essais de cristallisation 2D (clichés de Patrick Bron, CBS) Dans ces images, il faut chercher à repérer des arrangements réguliers, caractéristiques de l'organisation cristalline.

3.2.5.2. Analyses et conclusions

Le champ de l'observation directe de CF IA étant totalement inexploré, nous nous sommes penchés sur l'approche de microscopie électronique avec beaucoup d'ambitions. Le principal objectif est la détermination d'une enveloppe à basse résolution. Cependant, aucune des nombreuses conditions testées n'a permis l'acquisition d'images avec des objets distinguables, aussi bien en coloration négative qu'en Cryo-EM. Une stabilisation des facteurs par la méthode du GraFix n'a pas non plus permis d'observer des objets sur les clichés. Une exception doit être signalée pour Rna14p-Rna15p pour lequel les objets présents présentent

trop d'hétérogénéité pour permettre un traitement des images. Afin de nous affranchir des conditions de transport des échantillons, j'ai effectué une mission d'une semaine me permettant de tester sans plus de succès des échantillons non congelés. Ce déplacement a également été l'occasion de tester une approche originale et innovante, la cristallisation 2D, mais ces essais n'ont rien donné non plus.

Nous ne sommes pas capable d'expliquer de façon claire pourquoi nous ne sommes pas parvenus à prendre des clichés satisfaisants, mais comme lors de l'approche SAXS, la mobilité intrinsèque de CF IA semble le rendre réfractaire aux analyses.

3.2.1. Spectrométrie de masse en conditions natives

3.2.1.1. Principe

L'ensemble des résultats obtenus en spectrométrie de masse en conditions natives sont au crédit de la collaboration menée avec Sarah Cianferini-Sanglier et Johann Stojko au LCMS de Strasbourg.

Cette collaboration a été initiée avec pour principal objectif de répondre à la question de la stœchiométrie au sein de CF IA. Historiquement, si l'on écarte les cribles généraux (Gavin, Bosche et al. 2002, Krogan, Cagney et al. 2006), CstF-77 (homologue de Rna14p) est la seule séquence sur laquelle des essais de spectrométrie de masse aient été effectués (Bai, Auperin et al. 2007, Bai, Auperin et al. 2007) afin d'identifier des fragments générés par protéolyse ménagée pour la cristallisation. Nous pouvons donc encore une fois considérer comme vierge le champ de l'analyse de CF IA par spectrométrie de masse, *a fortiori* en conditions natives.

La spectrométrie de masse est une technique qui permet de détecter le ratio masse/charge et l'abondance d'ions dans une phase gazeuse. L'ionisation des protéines d'intérêt est réalisée grâce à une très fine dispersion de l'échantillon en solution. Le temps de vol de ces ions à travers l'appareil et ses différents équipements sera alors mesuré (Figure 45). Des propriétés comme la taille, la forme et la charge affectant les ions influence le temps de vol, les ions les plus volumineux ayant par exemple plus de collision avec le gaz et seront donc plus retardés que les petits ions.

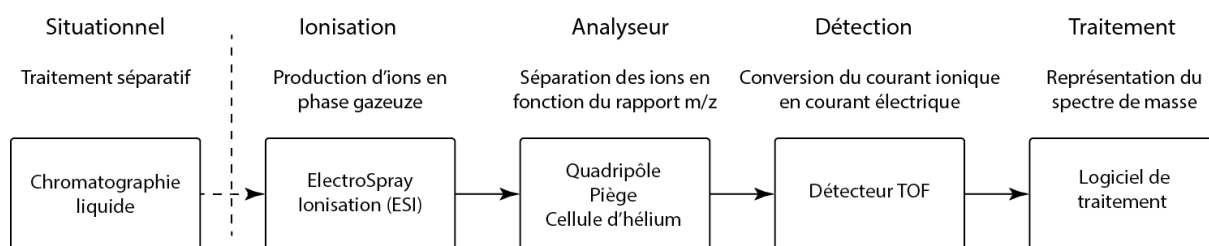


Figure 45 : Principe du spectroscope de masse.

Il s'agit d'une technique extrêmement sensible capable de détecter n'importe quelle espèce en solution à partir du moment où cette espèce génère des ions. Dans le cadre de l'analyse de complexes en conditions dénaturantes, il peut s'agir là d'une limite importante, où le phénomène de compétition à l'ionisation entre les partenaires peut empêcher la détection de certains d'entre eux. De la même manière, le moindre contaminant présent, s'il est ionisé, sera détecté. Pour faciliter l'interprétation des spectres déjà complexes générés par des facteurs multi-protéiques, il est donc impératif de maîtriser parfaitement la purification.

Dans le cas de la spectrométrie de masse en conditions natives, il faut également conserver les interactions non covalentes pour ne pas détruire notre complexe. Ceci est fait grâce à un processus d'ionisation douce à pression atmosphérique. Pour optimiser l'étape d'ionisation, les sels non volatiles comme les tampons à base de phosphates doivent être évités et remplacés par des solutions volatiles. L'acétate d'ammonium aqueux étant typiquement le meilleur compromis. Ils augmentent la force ionique de la solution et ne causent pas d'élargissement des pics. Une optimisation empirique des paramètres expérimentaux est également indispensable afin de conserver le complexe intact pendant son vol à travers l'instrument.

3.2.1.2. Analyses préliminaires

Afin de faciliter l'analyse des spectres de CF IA, nous avons commencé par caractériser les protéines seules dans la mesure du possible, ainsi que les sous complexes.

3.2.1.2.1. Rna14p et Rna15p

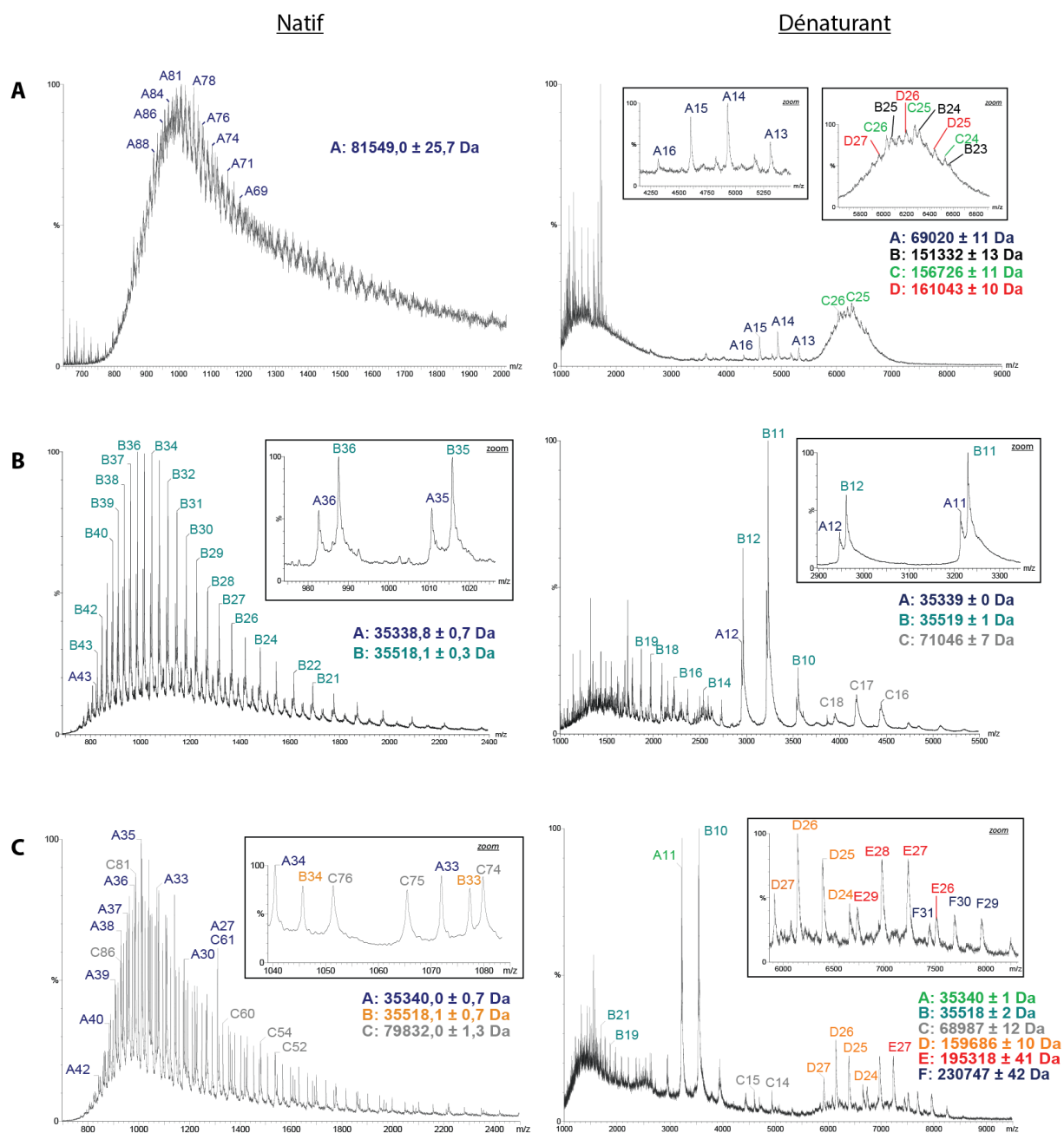


Figure 46 : Spectres de masse obtenus pour les échantillons de His-Rna14p (A), His-Rna15p (B) et His-Rna15p-Rna14p (C)

La Figure 46 présente ainsi dans le panneau A, spectre de gauche, le spectre obtenu avec l'analyse de His-Rna14p en condition dénaturantes. Rna14p est le premier échantillon que nous avons analysé, et nous pouvons voir que le dessalage n'est pas optimal, ce qui donne un bruit de fond élevé sur le spectre. Néanmoins, une espèce majoritaire unique est détectée, avec une masse molaire calculée de 81 549 Da (Tableau 4).

Espèces	Masse mesurée	Protéine assignée	Masse théorique (Da)	Différence de masse	Explications
---------	---------------	-------------------	----------------------	---------------------	--------------

	(Da)			(Da)	
A	81 549	His-Rna14p	82 668,2	- 1 119,2	Dégradation

Tableau 4 : *Espèce détectée dans l'analyse de His-Rna14p en conditions dénaturantes.*

Si l'on compare cette masse expérimentale à la masse théorique, on peut voir que la différence est de - 1 119,2 Da. Cela peut correspondre à la perte d'une dizaine de résidus environ. Nous savons par expérience que Rna14p se dégrade lorsqu'il est exprimé seul, notamment dans la partie C-terminale qui n'est pas structurée en absence de Rna15p. Si nous retirons les 11 résidus C-terminaux (EIPTVESTKSG), nous obtenons alors une masse théorique de 81 539 Da, dont la différence (10 Da) avec la masse mesurée est inférieure à la marge d'erreur de la mesure (25,7 Da).

Nous avons ensuite analysé l'échantillon en conditions natives (Figure 46, panneau A, à droite). Dans ces conditions, nous avons pu détecter une espèce dans les petits m/z et trois espèces avec des m/z plus importants (Tableau 5). Nous pouvons voir à nouveau que le dessalage non optimal entraîne un fort bruit de fond.

Espèces	Masse mesurée (Da)	Protéine assignée	Masse théorique (Da)	Différence de masse (Da)	Explications
A	69 020	His-Rna14p	82 668,2	- 13 648,2	Dégradation
B	151 332	His-Rna14p ₂	165 336,4	- 14 004,4	Dégradation
C	156 726	His-Rna14p ₂	165 336,4	- 8 610,4	Dégradation
D	161 043	His-Rna14p ₂	165 336,4	- 4 293,4	Dégradation

Tableau 5 : *Espèces détectées dans l'analyse de His-Rna14p en conditions natives.*

A nouveau, l'espèce la plus légère peut-être assignée à un monomère de His-Rna14p, mais la dégradation est ici plus importante que celle détectée en conditions dénaturantes. Cela est difficile à justifier avec exactitude, mais nous pouvons supposer que la différence de temps entre les deux analyses et le stockage de la protéine à 4°C pendant ce laps de temps est à l'origine de ce comportement. Les trois autres espèces détectées sont de masses relativement comparables, entre 151 kDa et 161 kDa. Elles peuvent toutes être assignées à des associations de monomères plus ou moins dégradés formant un dimère de His-Rna14, dont la masse théorique est de 165 kDa.

De ces analyses de Rna14p, nous confirmons le fait que Rna14p est peu stable quand il est exprimé seul. La dimérisation ne prévient pas la dégradation, laissant supposer qu'elle a principalement lieu au niveau de l'extrémité C-terminale. Malgré ces dégradations et un dessalage à optimiser, nous avons néanmoins réussi à caractériser les espèces produites par Rna14p seul.

La Figure 46 présente également, dans le panneau B, les analyses menées sur Rna15p en conditions dénaturantes (à gauche), ou natives (à droite). Nous pouvons remarquer que le dessalage est déjà meilleur. Dans le spectre en conditions dénaturantes, deux espèces sont identifiées (Tableau 6).

Espèces	Masse mesurée (Da)	Protéine assignée	Masse théorique (Da)	Différence de masse (Da)	Explications
A	35 338,8	His-Rna15p	35 498,5	- 159,7	Dégradation
B	35 518,1	His-Rna15p	35 498,5	+ 18,3	Variant de séquence Modification

Tableau 6 : Espèces détectées dans l'analyse de His-Rna15p en conditions dénaturantes.

Les deux espèces sont d'une masse très proche de la masse théorique. L'espèce B notamment, avec seulement 18,3 Da de plus, laisse penser soit à un variant de séquence, soit à une légère modification post-traductionnelle, comme une oxydation. Un séquençage du plasmide révèle finalement une mutation V196A (GTT -> GCT) dans la séquence du plasmide utilisé pour l'expression. Ce résidu, situé juste à la sortie du domaine « hinge » d'interaction avec Rna14p n'est partagé qu'avec *Kluyveromyces lactis* durant l'évolution. Le Tableau 7 présente l'analyse des espèces détectées à la lumière de cette mutation.

Espèces	Masse mesurée (Da)	Protéine assignée	Masse théorique (Da)	Différence de masse (Da)	Explications
A	35 338,8	His-Rna15p	35 470,4	- 131,6	Dégradation : perte de la méthionine N-terminale
B	35 518,1	His-Rna15p	35 470,4	+ 47,7	Modification : acétylation de la méthionine N-terminale

Tableau 7 : Correction des espèces détectées dans l'analyse de His-Rna15p en conditions dénaturantes.

Il faut remarquer que si le changement n'est pas très important, la différence de masse entre l'espèce A et la masse théorique est maintenant de - 131,6 Da, ce qui évoque la perte

de la méthionine N-terminale (131,2 Da). L'espèce B, la plus représentée, observe quant à elle une différence de 47,7 Da. Très certainement la marque de modification post-traductionnelle, il est délicat de statuer avec exactitude. Nous pouvons cependant penser qu'il s'agit d'une acétylation (42 Da) de la méthionine N-terminale, car c'est une modification relativement fréquente. Cette analyse en conditions dénaturantes nous révèle la présence de deux formes de His-Rna15p.

Si nous nous intéressons maintenant au spectre généré par l'analyse de l'échantillon en conditions natives, nous observons les espèces référencées dans le Tableau 8.

Espèces	Masse mesurée (Da)	Protéine assignée	Masse théorique (Da)	Différence de masse (Da)	Explications
A	35 339	His-Rna15p	35 470,4	- 131,4	Dégradation : perte de la méthionine N-terminale
B	35 519	His-Rna15p	35 470,4	+ 48,6	Modification : acétylation de la méthionine N-terminale
C	71 046	His-Rna15p ₂	70 940,8	- 105,2	

Tableau 8 : *Espèces détectées dans l'analyse de His-Rna15p en conditions natives.*

Dans cette analyse, nous retrouvons 2 espèces majoritaires, A et B, qui correspondent respectivement aux deux formes A et B du monomère His-Rna15p que nous avons déjà détecté en conditions dénaturantes. Nous observons également une troisième espèce C dont la masse élevée (71 kDa) fait indubitablement penser à un dimère de His-Rna15p (71 kDa). La détection de cet analyte est tout à fait surprenante, car il n'a jamais été fait mention de la moindre dimérisation de Rna15p, et nos essais de co-purification indiquent clairement qu'aucune construction de Rna15p ne parvient à dimériser avec Rna15p entier (3.1.1.3). Cette espèce serait-elle alors un artefact de notre expérience, créée par exemple à cause de la concentration ? Car même lors de la purification de l'échantillon His-Rna15p, aucun pic n'est apparu avant le pic du monomère durant l'exclusion de taille. Pour néanmoins tester la réalité de cette association, nous avons tout d'abord fait des essais de variation de voltage au sein du spectromètre, l'espèce C se montrant résistante à ce traitement, indiquant une association non artéfactuelle. Cependant, des essais de pull-down seront incapables de faire venir une copie non-tagguée avec la copie tagguée (cf 3.1.1.3 p.39). Il est donc possible que cette espèce soit exclusivement générée lors de la concentration du facteur.

L'analyse de His-Rna15p seul nous aura donc permis de bien caractériser son comportement et notamment de constater sa très grande stabilité. Nous avons également sensiblement amélioré le dessalage des échantillons.

Les deux protéines Rna14p et Rna15p ayant été caractérisées individuellement, nous nous sommes alors intéressés au sous-complexe His-Rna15p-Rna14p, dont les résultats sont présentés en Figure 46, panneau C. Le Tableau 9 présente les espèces retrouvées durant l'analyse en conditions dénaturantes.

Espèces	Masse mesurée (Da)	Protéine assignée	Masse théorique (Da)	Différence de masse (Da)	Explications
A	35 340	His-Rna15p	35 470,4	- 130,4	Dégradation : perte de la méthionine N-terminale
B	35 518	His-Rna15p	35 470,4	+ 47,6	Modification : acétylation de la méthionine N-terminale
C	79 832	Rna14p	79 960,4	- 128	Dégradation Variant de séquence

Tableau 9 : *Espèces détectées dans l'analyse de His-Rna15p-Rna14p en conditions dénaturantes.*

Nous retrouvons avec la même exactitude les espèces A et B générées par l'hétérogénéité dans les monomères de His-Rna15p. Une nouvelle espèce apparaît ici, qui est très proche de la masse attendue pour le monomère de Rna14p ($\Delta = -128$ Da). Contrairement à ce que nous observions avec Rna14p exprimé seul, nous avons maintenant un signal homogène et n'observons plus la moindre dégradation dans cet échantillon, ce qui confirme l'expérience que nous avons avec cette protéine. L'interaction entre le domaine central de Rna15p et le C-terminal de Rna14p prévient la dégradation de ce dernier. Grâce à cette analyse en conditions dénaturante, nous voyons que nous sommes capables de détecter tous les partenaires présents dans le sous-complexe et de leur attribuer une masse exacte, ce qui nous ouvre la voie pour une analyse en conditions natives.

Le Tableau 10 recense les six espèces assignées au spectre enregistré avec l'échantillon His-Rna15p-Rna14p en conditions natives. Il faut remarquer la détection de signal à partir de 7000 m/z qui est absent des spectres des protéines individuelles, et qui pourrait correspondre à la formation de complexes non covalents entre les partenaires. Si nous nous penchons sur le détail des espèces identifiées, nous voyons que nous retrouvons les espèces A et B, associées à Rna15p, ainsi que l'espèce C, pouvant correspondre à un produit de dégradation

de Rna14p. Nous retrouvons également dans l'espèce D l'homodimère de Rna14p. Les deux autres espèces, responsables des pics détectés au-delà de 7000 m/z, sont, elles, inédites.

Espèces	Masse mesurée (Da)	Protéine assignée	Masse théorique (Da)	Différence de masse (Da)	Explications
A	35 340	His-Rna15p	35 470,4	- 130,4	Dégradation : perte de la méthionine N-terminale
B	35 518	His-Rna15p	35 470,4	+ 47,6	Modification : acétylation de la méthionine N-terminale
C	68 987	Rna14p	79 960,4	- 10 973,4	Dégradation
D	159 686	Rna14p ₂	159 920,8	- 234,8	
E	195 318	His-Rna15p-Rna14p ₂	195 391,2	- 73,2	
F	230 747	His-Rna15p ₂ -Rna14p ₂	230 861,6	- 114,6	

Tableau 10 : *Espèces détectées dans l'analyse de His-Rna15p-Rna14p en conditions natives.*

L'espèce F, mesurée à 230 kDa, est très proche de la masse théorique de l'hétérotétramère Rna14p₂-Rna15p₂ décrit dans la littérature. Et si l'on considère un complexe formé par un objet D et deux objets B, on obtient un complexe de 230 722 Da, soit à 25 Da près, la masse de l'espèce F.

L'espèce E, à 195 kDa, se situe entre les masses attribuées au dimère de Rna14p et à l'hétérotétramère Rna14p₂-Rna15p₂. En suivant cette logique, nous avons essayé de l'assigner à un complexe formé par deux Rna14p et un seul Rna15p. Avec succès puisque ce complexe sera alors 73 Da plus léger qu'un complexe formé avec des protéines complètes et 114 Da plus lourd qu'un complexe formé par un objet D et un objet B.

Cette analyse de l'échantillon de Rna14p et Rna15p recombinants en conditions natives nous permet de bien identifier plusieurs populations. Les protéines seules, où nous avons pu remarquer la constance des deux formes de Rna15p et la tendance à la dégradation de Rna14p. Les complexes, parmi lesquels le dimère de Rna14p et l'hétérotétramère de stœchiométrie 1 :1, tous deux plusieurs fois décrits dans la littérature. Le troisième complexe, le trimère, n'a pour sa part jamais été décrit. Les pics de cette population sont bien dessinés, et plus importants que ceux de l'hétérotétramère, suggérant une plus grande population de ces ions. Cette espèce pourrait être un intermédiaire entre le dimère et le tétramère, ou

encore un artefact créé par les conditions. Car nous devons également remarquer que le dimère n'est pas dégradé, contrairement à ce que nous observions quand Rna14p était exprimé seul. Comme nous savons que Rna15p protège Rna14p de la dégradation, cela pourrait signifier que nous avons une association/dissociation des différents éléments, peut-être favorisée par les conditions expérimentales.

3.2.1.2.2. Pcf11p et Clp1p

Nous nous sommes ensuite intéressés aux deux autres composants de CF IA que sont Pcf11p et Clp1p, dont les spectres sont présentés en Figure 47. De notre expérience au laboratoire, Pcf11p et Clp1p ne sont pas stables quand ils sont exprimés seuls. Par contre, si une coexpression permet de stabiliser Clp1p, Pcf11p entier reste sujet à la dégradation. Pour éviter ceci, nous avons exprimé exclusivement le fragment d'interaction avec Clp1p, la construction allant du résidu 454 au résidu 563.

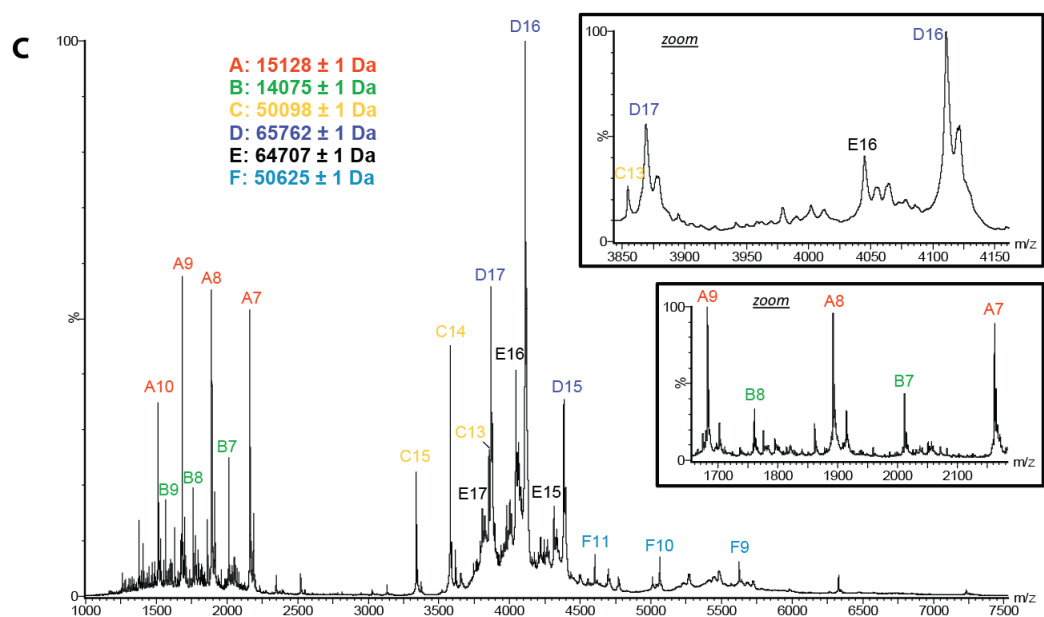
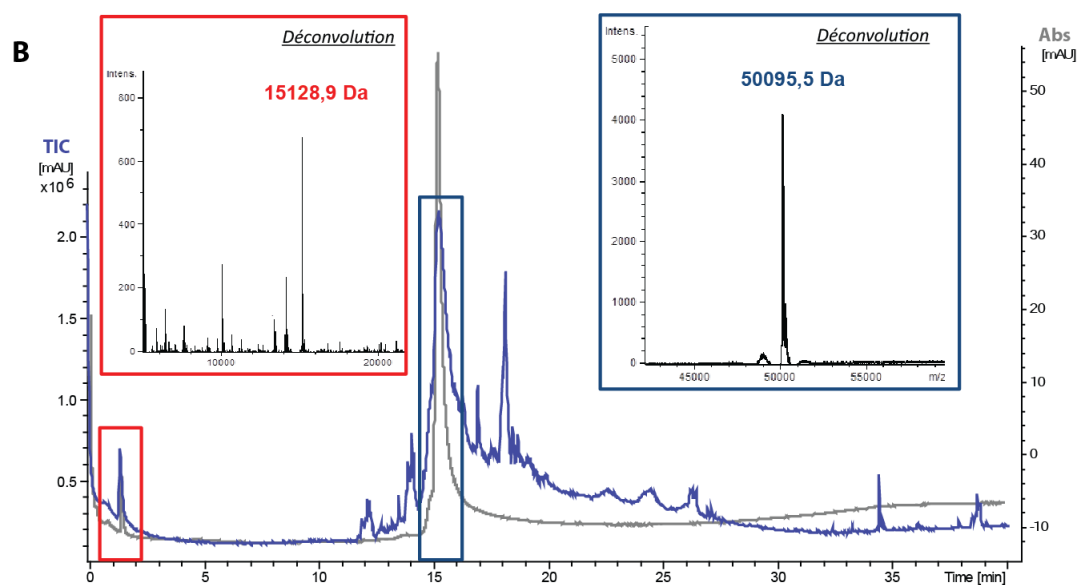
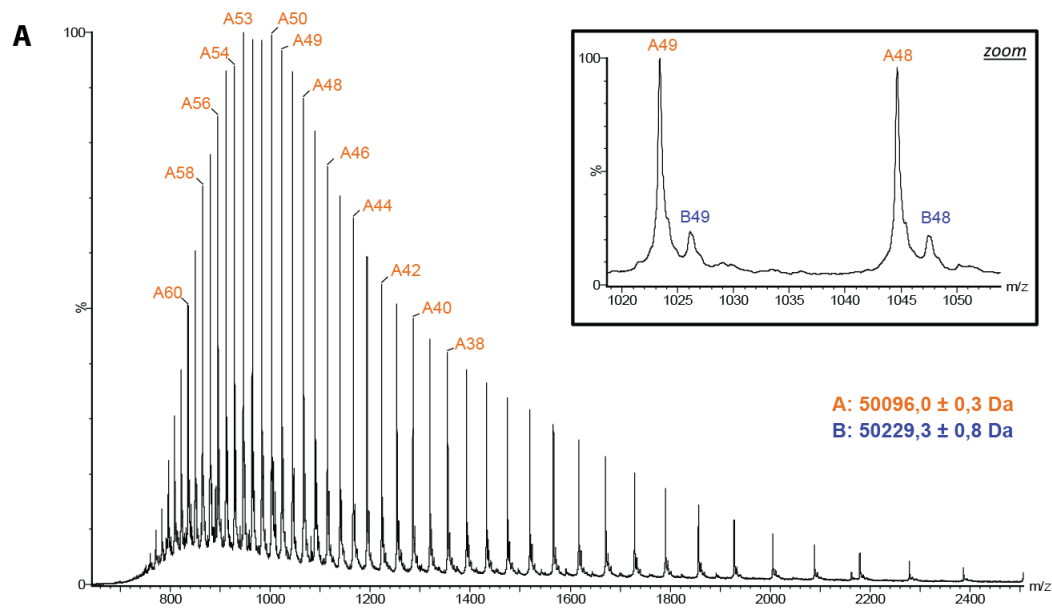
Le panneau A représente le spectre obtenu en condition dénaturantes. Deux espèces, reprises dans le Tableau 11, y sont détectées.

Espèces	Masse mesurée (Da)	Protéine assignée	Masse théorique (Da)	Différence de masse (Da)	Explications
A	50 096	Clp1p	50 225,7	- 129,7	Dégradation : perte de la méthionine N-terminale
B	50 229,3	Clp1p	50 225,7	+ 3,6	

Tableau 11 : *Espèces détectées dans l'analyse de His-Pcf11p(454-563)-Clp1p en conditions dénaturantes.*

La seconde espèce, B, largement minoritaire, correspond presque exactement à la masse théorique de Clp1p. L'espèce A correspond à une perte de 133,3 Da par rapport à l'espèce B, ce qui peut être interprété comme la perte de la méthionine N-terminale. Il faut alors remarquer qu'aucune espèce n'est attribuée au fragment de Pcf11p attendu à 15 kDa. Du fait de la très forte compétition à l'ionisation, nous avons décidé d'analyser cet échantillon par LC-MS.

Figure 47 : *Spectres obtenus en conditions dénaturantes (A), LC-MS (B) ou natives (C) à partir de l'échantillon His-Pcf11p (454-563)-Clp1p*



Les spectres sont présentés dans le panneau B, et nous pouvons voir que nous observons bien les deux pics attendus. Le Tableau 12 présente les informations sur les espèces identifiées à la déconvolution de ces deux pics.

Espèces	Masse mesurée (Da)	Protéine assignée	Masse théorique (Da)	Différence de masse (Da)	Explications
A	15 128,9	His-Pcf11p (454-563)	15 276,2	- 147,3	Dégradation : perte de la méthionine N-terminale Oxydation
B	50 095,5	Clp1p	50 225,7	- 130,2	Dégradation : perte de la méthionine N-terminale

Tableau 12 : *Espèces détectées dans l'analyse de His-Pcf11p(454-563)-Clp1p en conditions dénaturantes précédée par un chromatographie en phase liquide (LC-MS). La courbe grise représente l'absorbance à 280nm. La courbe violette représente le courant ionique total détecté.*

La seconde espèce correspond à l'espèce majoritaire de Clp1p qui avait accaparé l'ionisation dans l'essai en conditions dénaturantes. La première espèce cependant correspond à une masse de 15 kDa, soit uniquement 147 Da de moins que la masse théorique attendue pour His-Pcf11p (454-563). Cette différence de 147 Da pourrait s'expliquer par la perte de la méthionine terminale (131 Da) et une oxydation (16 Da).

Nous deux composants étant maintenant caractérisés, nous avons analysé le complexe en conditions natives. Le spectre est présenté dans le panneau C et les espèces identifiées dans le Tableau 13.

Un grand nombre d'espèces sont détectées. La A et la C ont été identifiées précédemment, et la B semble être un variant de dégradation de His-Pcf11p (454-563).

L'espèce D fait 260,1 Da de plus que la masse théorique du dimère. Cependant, si nous formons un dimère avec les espèces A et C, nous obtenons maintenant une différence de 536 Da entre la masse du complexe formé par les protéines et la masse enregistrée. Nous savons que Clp1p interagit avec l'ATP (507 Da) et un ion Mg^{2+} (23 Da) et ce sont certainement eux que nous observons ici. L'espèce E, de masse légèrement moins importante, doit également être constituée par une association de ce type, mais ayant subi une légère dégradation.

Espèces	Masse mesurée (Da)	Protéine assignée	Masse théorique (Da)	Différence de masse (Da)	Explications
A	15 128	His-Pcf11p (454-563)	15 276,2	- 148,2	Dégradation : perte de la méthionine N-terminale Oxydation
B	14 075	His-Pcf11p (454-563)	15 276,2	- 1 201,2	Dégradation
C	50 098	Clp1p	50 225,7	- 127,7	Dégradation : perte de la méthionine N-terminale
D	65 762	His-Pcf11p (454-563)- Cpl1p	65 501,9	+ 260,1	Dégradation : perte de la méthionine N-terminale sur les deux protéines Oxydation de Pcf11p Cofacteurs : ATP et Mg ²⁺
E	64 707	His-Pcf11p (454-563)- Cpl1p	65 501,9	- 794,9	Dégradation
F	50 625	Clp1p	50 225,7	+ 399,3	Dégradation : perte de la méthionine N-terminale Cofacteurs : ATP et Mg ²⁺

Tableau 13 : *Espèces détectées dans l'analyse de His-Pcf11p(454-563)-Clp1p en conditions natives.*

Le même raisonnement peut être appliqué à l'espèce F qui fait 399,3 Da de plus que la masse théorique de Clp1p, mais 527 Da de plus que l'espèce C. L'espèce F serait donc formée par un Clp1p et ses ligands ATP et Mg²⁺.

Même si nous n'avons pas pu travailler sur la caractérisation de Pcf11p entier, nous avons réussi à déterminer le comportement de Clp1p en conditions dénaturantes et natives. Nous avons notamment pu identifier avec précision son interaction avec une molécule d'ATP et un atome de magnésium.

3.2.1.3. CF IA : le complexe entier

Après cette caractérisation des partenaires individuels et des sous-complexes, nous avons analysé le complexe CF IA. Ce complexe est issu de la coexpression et copurification des quatre sous-unités. La construction de Pcf11p utilisée est le His-Pcf11p ΔQ.

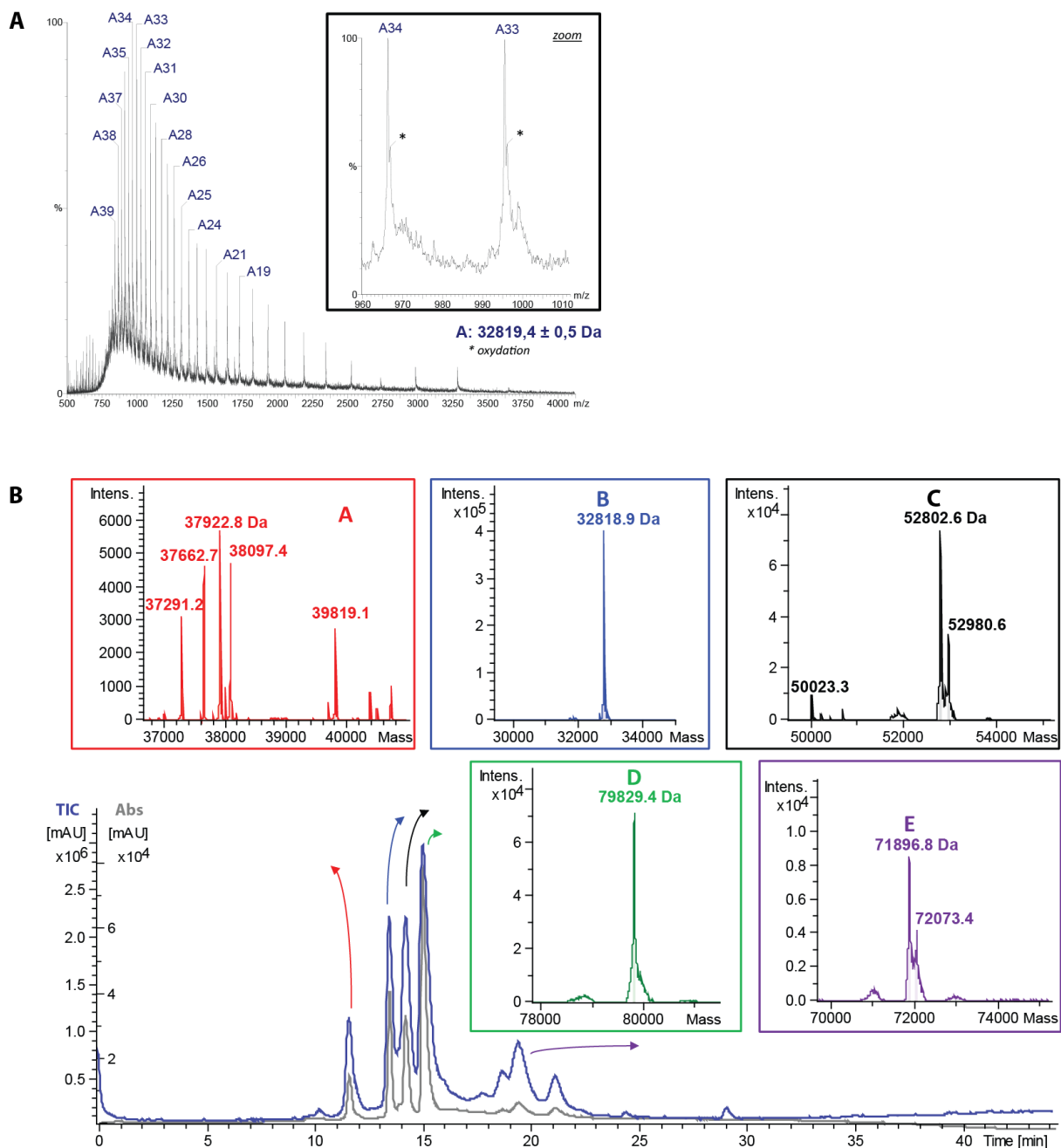


Figure 48 : A : Résultats obtenus en conditions dénaturantes pour CF IA His-Pcf11p Δ Q. B l'analyse est précédée d'une séparation en phase liquide (LC-MS)

Le spectre obtenu en conditions dénaturantes est présenté en panneau A de la Figure 48. Le Tableau 14 résume l'espèce qui est identifiée.

Espèces	Masse mesurée (Da)	Protéine assignée	Masse théorique (Da)	Différence de masse (Da)	Explications
A	32 819,4	Rna15p	32 762,6	+ 56,8	Modification Variant de séquence

Tableau 14 : Espèces détectée dans l'analyse de CF IA His-Pcf11p Δ Q en conditions dénaturantes.

Nous pouvons voir qu'il y a une forte compétition à l'ionisation au sein du complexe et que seule une espèce assignée comme Rna15p produit des ions.

Pour nous affranchir de cela, nous avons étudié le complexe en LC-MS.

Espèces	Masse mesurée (Da)	Protéine assignée	Masse théorique (Da)	Différence de masse (Da)	Explications
A	>37 291 <39 819	His-Pcf11p ΔQ	72 043,3		Dégradation
B	32 818,9	Rna15p	32 762,6	+ 56,3	Modification Variant de séquence
C	52 802,6	His-Clp1p	52 933,5	- 130,9	Dégradation : perte de la méthionine N-terminale
D	79 829,4	Rna14p	79 960,4	- 131	Dégradation : perte de la méthionine N-terminale
E	71 896,8	His-Pcf11p ΔQ	72 043,3	- 146,5	Dégradation : perte de la méthionine N-terminale Oxydation

Tableau 15 : Espèces détectée dans l'analyse de CF IA His-Pcf11p ΔQ en conditions dénaturantes précédée par une chromatographie en phase liquide (LC-MS). La courbe grise représente l'absorbance à 280nm. La courbe violette représente le courant ionique total détecté.

Les quatre protéines sont facilement identifiables. Comme nous nous y attendions, Pcf11p se dégrade, mais nous sommes tout de même capable de visualiser une espèce de Pcf11p non dégradée.

Nous avons ensuite analysé le complexe en conditions natives (Figure 49). Les panneaux A et B représentent les spectres obtenus avec différents niveaux d'énergie dans l'appareil, et nous pouvons immédiatement remarquer que la désolvatation est bien meilleure avec une énergie plus importante.

Neuf espèces sont identifiées sur le spectre présenté dans le panneau B et l'agrandissement du panneau C. Ces espèces sont détaillées dans le Tableau 16 et les pages suivantes. Il sera peut-être plus confortable de lire les pages suivantes en ayant le schéma de la Figure 50, page 115, sous les yeux.

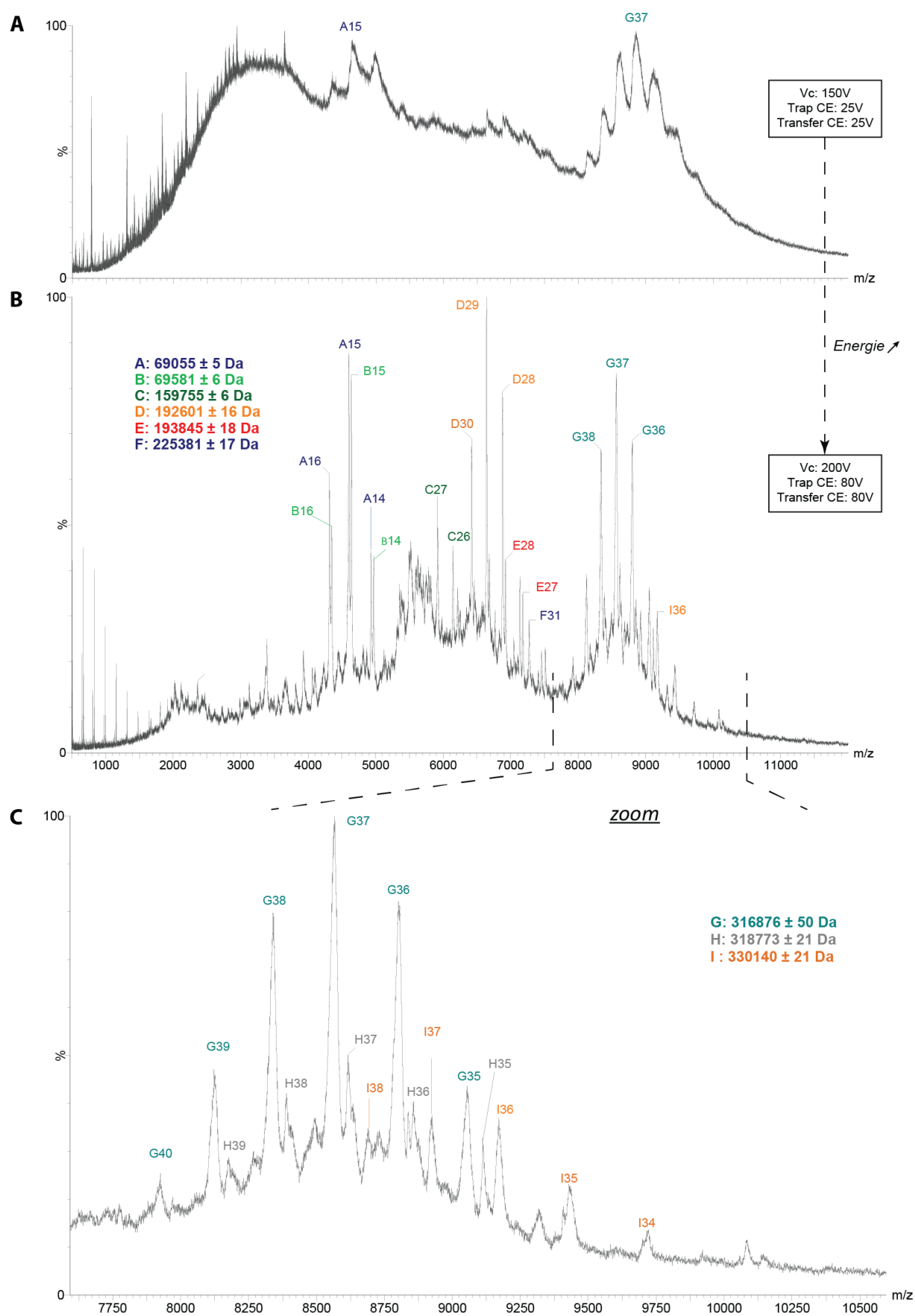


Figure 49 : Analyse du complexe CF IA His-Pcf11 Δ Q en conditions natives à faible énergie (A) ou à haute énergie (B et agrandissement des hauts m/z en C)

Espèces	Masse	Explications		
A	69 055	Dégradation Pcf11p ^{dénat}		
B	69 581	Dégradation Pcf11p ^{dénat}		
C	159 755	Dimérisation de Rna14p ^{dénat} + 96 Da		
D	192 601	C + Rna15p ^{dénat} + 27,1 Da		
E	193 845	D + fragment 1244 Da		
F	225 381	D + Rna15p ^{dénat} – 38,9 Da		
G	316 876	<u>Solution G1:</u> B + e + His-Clp1p ^{dénat} + 2 Zn ²⁺ + Mg ²⁺ + ATP -> 316 913,6 Da (Δ -37,6 Da) Stœchio.: 2:1:1:1	<u>Solution G2 :</u> 2 Rna14p ^{dénat} + Rna15p ^{dénat} + His-Clp1p ^{dénat} + His-Pcf11p ^{dénat} + 2 Zn ²⁺ -> 317307,1 Da (Δ -431,1 Da) Stœchio.: 2:1:1:1	<u>Solution G3 :</u> D+ Rna15p ^{dénat} + His-Clp1p ^{dénat} + His-Pcf11p (37 922,8 Da) + 2 Zn ²⁺ + Mg ²⁺ + ATP -> 316 830,3 Da (Δ -45,7 Da) Stœchio.: 2:2:1:1
H	318 773	<u>Solution H1 :</u> G1 avec Pcf11p ^{dénat} soit : His-Pcf11p ^{dénat} + D + His-Clp1p ^{dénat} + 2 Zn ²⁺ + Mg ²⁺ + ATP + frag ^{nt} 1244 Da -> 319 229 Da (Δ +456 Da) Stœchio.: 2:1:1:1	<u>Solution H2 :</u> G2 + Mg ²⁺ + ATP + frag ^{nt} 1244 Da soit : 2 Rna14p ^{dénat} + Rna15p ^{dénat} + His-Clp1p ^{dénat} + His-Pcf11p ^{dénat} + 2 Zn ²⁺ + Mg ²⁺ + ATP + frag ^{nt} 1244 Da -> 319106,1 (Δ -331,1 Da) Stœchio.: 2:1:1:1	<u>Solution indépendante H4 :</u> D + His-Clp1p ^{dénat} + His-Pcf11p ^{dénat} + 2 Zn ²⁺ + frag ^{nt} 1244 Da -> 318 674,4 Da (Δ -98,6 Da) Stœchio.: 2:1:1:1
I	330 140	<u>Solution I1 :</u> H + un large fragment (12 495,8 Da) -> 330 140 Da Stœchio.: 2:1:1:1		<u>Solution I3 :</u> G3 avec His-Pcf11p (51 379 Da) (espèce non observée) au lieu de His-Pcf11p (37 922,8 Da) -> 330 140 Da Stœchio.: 2:2:1:1

Tableau 16 : Espèces détectées dans l'analyse de CF IA His-Pcf11p ΔQ en conditions natives. Une version schématique est proposée dans la Figure 50.

Non observées en conditions dénaturantes, les espèces A et B, respectivement de 69 et 69,5 kDa, sont très certainement des produits de dégradation. Pcf11p étant de loin la protéine la moins stable, ces fragments peuvent être attribués à du Pcf11p dégradé. L'espèce C est pour sa part légèrement plus massive que la masse attendue pour une dimérisation du Rna14p qui aurait perdu sa méthionine N-terminale.

L'espèce D, extrêmement représentée, peut être très correctement reconstituée à partir de l'espèce C et d'une molécule de Rna15p. Cette espèce correspond alors à un trimère formé par Rna14p et Rna15p en stœchiométrie 2 :1.

Les espèces E et F peuvent alors être assignées à partir de l'espèce D. La différence entre C et E est de 1244 Da, ce qui ne s'approche de rien en particulier, et que nous avons donc choisi d'assigner à un fragment de dégradation car nous le retrouvons dans certaines solutions proposées pour les espèces G et H. L'espèce F peut pour sa part être parfaitement retrouvée en ajoutant une copie de Rna15p à l'espèce D. Nous nous retrouvons alors en présence d'un hétérotétramère de stœchiométrie 2 :2.

Les espèces G H et I, avec respectivement 317, 319 et 330 kDa, sont indéniablement des formes de CF IA. De par leurs tailles et le nombre de composants, plusieurs solutions peuvent être proposées.

Dans la solution 1, l'espèce G (G1) est constituée comme une suite logique aux espèces B, D, E et F observées en conditions natives. Elle est composée de l'espèce D, l'espèce B, le His-Clp1p identifié en conditions dénaturantes, la totalité des cofacteurs connus (2 atomes de Zinc, un ATP et un magnésium) et complété par le fragment de 1244 Da identifié précédemment. Cette entité observe alors une stœchiométrie 2 :1 :1 :1 et une masse égale à celle mesurée. La différence est en effet de 37,6 Da, mais l'imprécision sur la mesure est de 50 Da. L'espèce H1 serait en tout point similaire à l'espèce G1, si ce n'est que sa copie de Pcf11p ne serait pas dégradée mais correspondrait à la masse enregistrée en condition dénaturantes. Cette solution H1 observe cependant une différence non négligeable de 456 Da avec la masse mesurée pour l'espèce H.

Dans cette solution, l'attribution de G1 est très solide. La différence entre les masses théoriques et mesurées est inférieure aux imprécisions liées aux conditions expérimentales.

G1 et H1 sont constituées d'espèces précédemment identifiées, au premier lieu desquelles l'espèce D, qui est une espèce majoritaire sur le spectre. Cette solution permet également de construire un complexe composé de toutes les protéines observées en conditions dénaturantes. Cependant, la solution H1 est 1,4‰ différente de la masse attendue. De plus G1 et H1 intègrent le fragment de 1244 Da observé dans l'espèce E, mais que je ne suis pas capable d'expliquer. Globalement, cette solution est extrêmement solide sur l'attribution de G1, mais plus faible sur celle de H1.

Dans la solution 2, une stœchiométrie 2 :1 :1 :1 est également adoptée par le complexe qui est dans ce cas composé en intégralité des protéines observées en conditions dénaturantes. La solution G2 considère un complexe composé par les quatre protéines identifiées en conditions dénaturantes et deux atomes de zinc. Cette solution est alors 431,1 Da plus lourde que la masse mesurée. L'espèce H2 est alors constituée par G2 enrichie en ATP, magnésium et avec le fragment de 1244 Da. Ainsi constituée, G2 est 331,1 Da plus lourde que la masse expérimentale.

Cette solution possède deux points forts. Tout d'abord, les complexes sont constitués en intégralité par les protéines observées en dénaturant, et donc le complexe est composé de protéines entières. Ensuite, même si G2 et H2 sont différents respectivement de 1,4‰ et de 1‰ des masses des espèces mesurées, ils expliquent relativement bien la différence de masse entre G et H. La différence est en effet de 1 897 Da, et la différence entre G2 et H2 est de 1 799 Da. Ainsi, la différence de H2 avec H est principalement le reflet de la différence entre la solution G2 et G. Cette différence entre G et G2 est le principal point faible de cette solution. Nous pouvons également nous demander pourquoi Clp1p perdrait son contenu en ATP et en Magnésium, mais nous avons vu dans l'analyse de Clp1p avec Pcf11p que cela était possible. Il faut également noter que nous avons toujours la présence de ce fragment de 1244 Da, que nous ne pouvons toujours pas expliquer. Dans ce scénario, ni G2 ni H2 ne sont non plus reliés à une des sous espèces observées en condition dénaturantes. Cette solution gère finalement correctement la relation entre les espèces G et H mais manque de solidité sur l'attribution de G2.

Une autre attribution de H peut toutefois être proposée, qui peut être reliée aussi bien à G1 que G2. Dans cette hypothèse, H4 est constitué par D, les formes de His-Clp1 et His-

Pcf11p observées en conditions dénaturantes ainsi que deux atomes de Zinc et le fragment de 1244 Da. Alors constitué, H4 conserve une stœchiométrie 2 :1 :1 :1 et fait 98,6 Da de moins que l'espèce enregistrée, soit 0,3‰ de différence. Si on le relie à G1, nous restons dans le contexte d'un complexe formé autour de l'espèce D, mais avec une perte/acquisition de l'ATP et du magnésium et une dégradation de Pcf11p. Au contraire, si nous le relions à G2, nous restons dans une lignée de complexes qui n'intégreraient pas l'ATP et la différence de masse est principalement comblée par le fragment de 1244 Da.

La dernière solution est une solution organisée autour de la dégradation de Pcf11p au sein d'un complexe en stœchiométrie 2 :2 :1 :1. G3 est alors composée par l'espèce D, une seconde copie de Rnap15p et une copie de Clp1p. Pcf11p est alors un fragment de dégradation de 37 922,8 Da, comme identifié dans le panneau A, en LC-MS. Deux atomes de zinc sont également présents, avec un ion magnésium et une molécule d'ATP. Le complexe ainsi formé fait 316 830,3 Da, soit uniquement 45,7 Da de moins que la masse détectée, une différence en deçà de la marge d'erreur de l'expérience. A partir de cette solution G, une solution H3 très simple peut être proposée en remplaçant le fragment de Pcf11p de 37 922,8 Da par un fragment moins dégradé, également identifié en LC-MS. La différence entre les deux états de dégradation est de 1896,3 Da couvre parfaitement la différence de 1 897 Da entre les deux espèces G et H. La solution H3 ferait alors 318 726,6 Da, soit 46,4 Da de différence avec la masse expérimentale de H.

Cette solution présente l'avantage de se baser sur l'espèce D complétée par des protéines observées en conditions dénaturantes. Tous les cofacteurs sont présents et les masses obtenues sont extrêmement proches des masses mesurées. De la même manière, la différence entre les espèces G et H est expliquée de manière simple et élégante. Cependant, si cette solution a le mérite d'exploiter les produits de dégradation observés en condition dénaturantes, rien ne nous assure qu'il s'agisse effectivement de fragments de Pcf11p. Et alors que nous observons également du His-Pcf11p non dégradé en LC-MS, cette solution ne permet pas de l'utiliser dans la reconstitution d'un complexe.

Il reste une espèce détectée en conditions natives que nous n'avons pas encore abordée. L'espèce I, détectée à 330 kDa, est supérieure à la masse d'un complexe de type 2 :1 :1 :1 mais largement inférieure à celle d'un complexe de type 2 :2 :1 :1. De ce fait, aucune

solution convaincante ne peut réellement être proposée. Il peut en effet s'agir d'un complexe H issu de n'importe laquelle des quatre propositions, auquel s'est adjoint un fragment de 12,5 kDa. Pour rester dans la logique de la solution 3 autour de la dégradation de Pcf11p, il s'agirait alors d'un complexe G3 avec un Pcf11p de 51 379 Da, espèce que nous n'avons pas détectée durant les analyses en conditions dénaturantes.

Le Tableau 17 récapitule les points forts et points faibles des différentes solutions proposées, et un schéma récapitulatif est proposé en Figure 50.

Solution	Pro	Con
1	<ul style="list-style-type: none"> La masse obtenue pour G1 est très convaincante G1 et H1 sont basées sur d'autres espèces observées sur le spectre H1 est composé de protéines détectées en LC-MS 	<ul style="list-style-type: none"> Pas d'explication pour le fragment de 1244 Da La masse assignée à H1 n'est pas exacte
2	<ul style="list-style-type: none"> G2 et H2 sont composés des protéines détectées en LC-MS La différence entre G et H bien expliquée 	<ul style="list-style-type: none"> Les masses assignées à G2 et H2 sont un peu trop importantes Pas d'explication pour le fragment de 1244 Da
3	<ul style="list-style-type: none"> Les masses obtenues pour G3 et H3 sont très convaincantes La différence entre G et H bien expliquée Explication de certaines espèces dégradées observées en LC-MS 	<ul style="list-style-type: none"> Aucune espèce ne contient le Pcf11p entier détecté en LC-MS On ne sait pas si les fragments de dégradation sont issus de la dégradation de Pcf11p
4	<ul style="list-style-type: none"> Peut-être lié à G1 ou G2 	<ul style="list-style-type: none"> Absence totale d'ATP et de magnésium Pas d'explication pour le fragment de 1244 Da

Tableau 17 : *Tableau résumant les points forts et les points faibles des différentes solutions proposées dans le Tableau 16.*

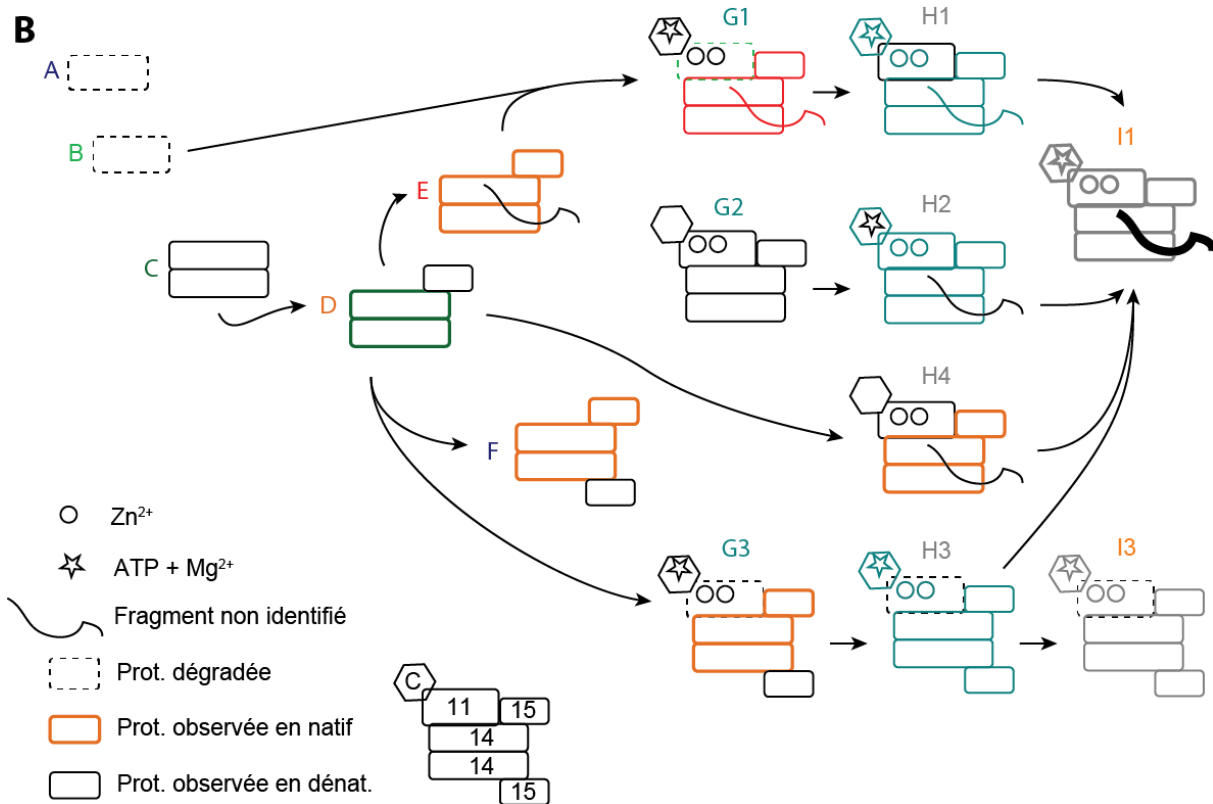
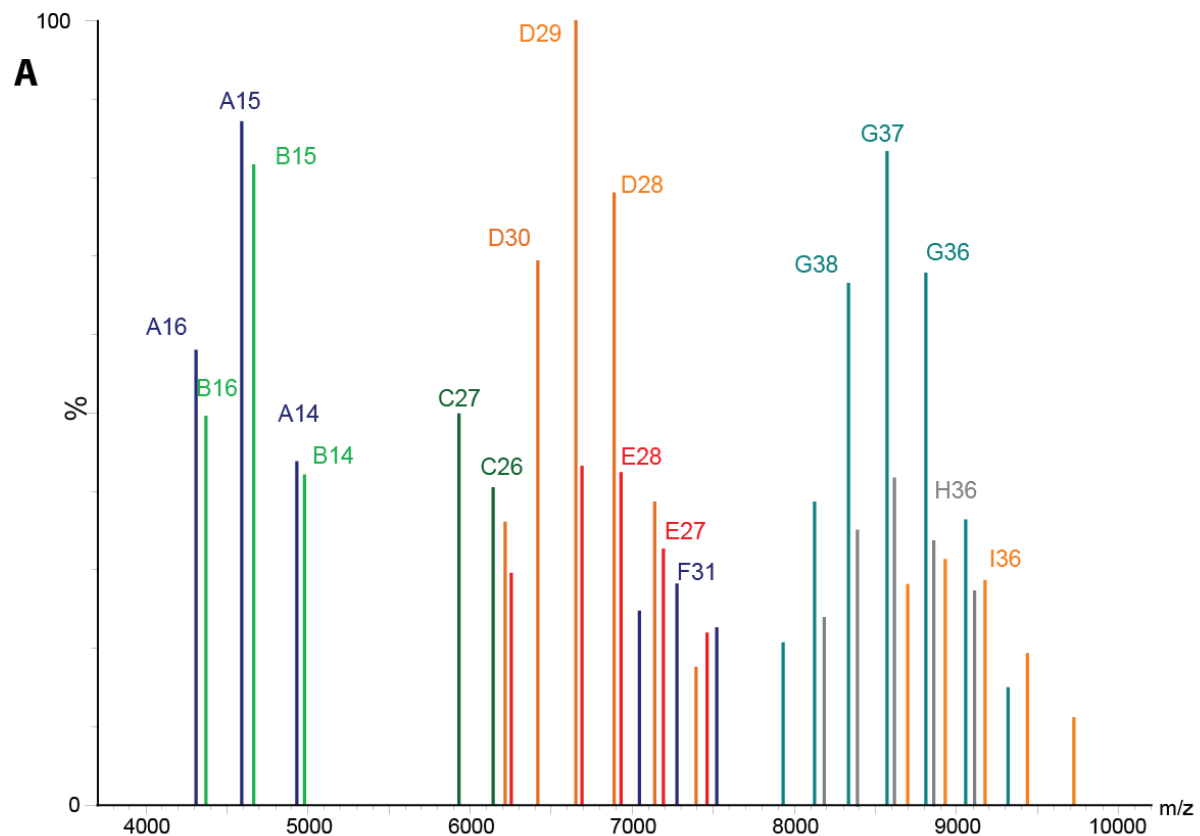


Figure 50 : Récapitulatif de l'analyse du CF IA en conditions natives. **A.** Schématisation du spectre obtenu en conditions natives. (Figure 49) **B.** Diagramme reprenant les propositions du Tableau 16 concernant l'attribution des espèces observées et les relations entre elles.

3.2.1.4. Conclusions

En se lançant dans une analyse du complexe en spectroscopie de masse en conditions natives, notre objectif est de répondre de manière directe et sans équivoque à la question de la stœchiométrie au sein de CF IA. Pour le moment, les mesures effectuées sur le complexe entier ne permettent pas de détecter des formes non dégradées qui nous auraient données une réponse claire. De ce fait, nous avons pour l'instant des hypothèses, et encore plus de questions. Pour y répondre de manière définitive, il sera nécessaire de refaire des analyses.

Si l'on regarde la littérature, on trouve des modèles de CF IA dans lequel le complexe est décrit avec une stœchiométrie 2 :2 :1 :1 (2 Rna14, 2 Rna15, 1 Pcf11 et 1 Clp1). Ce modèle n'est finalement basé sur aucune observation directe mais plutôt sur deux observations indépendantes. Le fait que Rna14 puisse dimériser, et le fait que Rna14 et Rna15 interagissent. Ces deux acquis, appuyés par le fait que les homologues de mammifère dimérisent également et que Pcf11-Cpl1 forment un complexe de stœchiométrie 1 :1, conduisent les auteurs de l'étude référence sur le sujet (Gordon, Shikov et al. 2011), à conclure sur une stœchiométrie 2 :2 :1 :1 pour le CF IA avec une analyse des intensités relative des bandes sur un gel au bleu de Coomassie comme principal argument. Cette démonstration a été acceptée, et la stœchiométrie du CF IA est depuis citée comme étant de cet ordre-là. La justification biologique se trouve résumée dans le fait que deux RRM valent mieux qu'un pour assurer le rôle de positionnement de la machinerie qu'est celui de CF IA.

Or, c'est également là que se pose une question, car la longueur classique du Positionning Element, extrêmement conservée de même que la séquence, est de 5 ou 6 nucléotides. Cependant, la zone d'interaction d'un RRM étant d'environ 4 bases, deux Rna15 devraient pas pouvoir interagir simultanément avec le PE. Deux hypothèses sont alors possibles. Il est proposé qu'un seul Rna15 interagisse avec le PE, et que sa spécificité lui soit apporté par la présence de Hrp1 (Valentini, Weiss et al. 1999, Gross and Moore 2001, Perez Canadillas and Varani 2003, Perez-Canadillas 2006, Leeper, Qu et al. 2010), tandis que le second Rna15 interagirait de manière non spécifique avec le pré-ARNm. On peut également imaginer que les deux RRM interagissent avec le PE, mais qu'une certaine dégénérescence de l'interaction soit tolérée sur les extrémités.

Mais il est également possible d'imaginer que Rna15 ne soit présent qu'en une seule copie au sein du CF IA. Nous voyons deux copies dans le complexe Rna14p-Rna15p analysé en conditions natives. Dans la structure qu'ils ont résolue, Paulson et Tong (Paulson and Tong 2012), ne voient qu'une seule région Hinge de Rna15kl par dimère de Rna14kl. Cela ne permet cependant pas de conclure quoi que ce soit, car il peut s'agir d'un simple biais d'agencement cristallin. L'existence de complexes CF IA alternatifs à par exemple déjà été proposée, dans lesquels Spt5 remplacerait Rna15p pour la maturation de certaines séquences (Lindstrom, Squazzo et al. 2003, Mayer, Schrieck et al. 2012). La possibilité de ne lier qu'une seule copie permettrait alors d'avoir des complexes homogènes, contenant soit l'une ou l'autre des sous-unité, mais jamais les deux.

Une asymétrie est créée au sein du CF IA. La question est : à quel moment, et quelles sont les sous-unités concernées. Dans le cas d'une Rna15p unique, lié via son CTD à Pcf11p, cela impliquerait très probablement l'interaction entre Rna14p et Pcf11p, qui préviendrait l'interaction avec un second Rna15p. Si nous avons deux Rna15p, nous avons alors deux sites potentiels d'interaction avec Pcf11p sur Rna15p. Il faut à ce moment-là que le site d'interaction avec Rna14p soit unique, afin d'exclure un potentiel second Pcf11p. Un site unique sur un dimère suggérerait qu'il soit formé lors de la dimérisation, hypothèse jamais clairement testée à ce jour à ma connaissance. Le mutant Rna14p ES (Gordon, Shikov et al. 2011), perturbant la dimérisation de Rna14p, est thermosensible *in vivo* et toujours capable de s'assembler en un CF IA « ES » dont le profil de migration sur gel filtration est identique au facteur sauvage, conduisant les auteurs à proposer que Pcf11p « pont » les deux Rna14p mutants. Mais cette possibilité est également viable avec une unique copie de Rna15p.

Une stratégie de purification de facteurs de levure exprimant deux constructions de Rna15p a été développée avec Julia Guéguénat, et devrait apporter des éléments de réponse, en parallèle avec de nouvelles analyses de masse.

3.3. Cristallisation de CF I

3.3.1. Quels partenaires pour stabiliser le CF IA

Une dernière piste à explorer pour essayer de contrôler la mobilité des domaines du CF IA consiste à essayer de le mettre en présence de ses cofacteurs. Dans notre cas, il s'agira de

CF IB (*i. e.* Hrp1), CPF et le pré-ARNm. Notre objectif étant de cependant de stabiliser notre objet, il nous faudra être vigilant, car ajouter des partenaires peut avoir finalement l'effet inverse. Afin d'éviter ce dernier cas de figure, et pour ne pas avoir un système trop complexe, nous avons décidé de ne travailler qu'avec les partenaires les mieux caractérisés que sont les RRM de Hrp1 et l'ARN (Kessler, Henry et al. 1997, Valentini, Weiss et al. 1999, Perez-Canadillas 2006, Leeper, Qu et al. 2010).

Nous pouvons espérer que les interactions engagées par CF IA avec ces partenaires vont limiter la mobilité de Rna15 et, dans une certaine mesure, de Rna14. L'ARN interagit exclusivement avec le RRM des Rna15 au sein du CF IA, et avec les RRM de Hrp1. Ces interactions sont extrêmement bien connues et une structure en complexe avec un ARN de synthèse a été résolue. Cette publication décrit par ailleurs une interaction qui s'établit entre les RRM de Hrp1 et celui de Rna15, via la proximité physique amenée par la liaison à l'ARN. C'est là que cette proposition est à nuancer, car les interactions ARN-protéines décrites dans la structure ne correspondent pas exactement à ce qui a pu être décrit par les études fonctionnelles. En effet, la structure a été obtenue avec un complexe constitué par un ARN de 13 nucléotides, le RRM de Rna15p et les RRM de Hrp1p. La proximité entre les RRM, et l'interaction directe qui est décrite, est donc en quelque sorte faussée par la petite taille de l'ARN. Une interaction a également été rapportée entre Rna14p et Hrp1p, mais rien de plus n'est décrit à ce sujet. En considérant toutes ces informations, nous pouvons espérer que l'ajout d'ARN et des RRM de Hrp1 organise les RRM des sous-unités Rna15 et apporte une interaction supplémentaire à Rna14.

Cependant, Si Rna15 est une source importante de mobilité, Pcf11 est également en grande partie inconnu et la région entre le CID et le domaine d'interaction avec Rna15-Rna14 est considérée comme non structurée. Comme nous avons décidé de ne pas travailler avec des partenaires du CPF, dont certains interagissent avec Pcf11, la solution la plus simple pour limiter l'impact de cette région est de la supprimer. Nous travaillerons donc pour ces essais avec un Pcf11 dit « minimal », caractérisé antérieurement dans cette thèse. Cette construction s'étend du résidu 322 au résidu 518, ce qui couvre les régions d'interaction avec Rna14kl-Rna15kl et Clp1kl.

La dernière décision à prendre vis-à-vis des partenaires concerne l'ARN. D'un point de vue pratique, je disposais de la séquence de Cyc1 de *S. cerevisiae*, mais j'avais pris la décision de travailler avec des facteurs de *K. lactis*. Cela ne sera cependant pas un problème car une étude récente (Moqtaderi, Geisberg et al. 2013) s'intéressant aux spécificités de la polyadénylation chez trois espèces de levures (*Saccharomyces cerevisiae*, *Kluyveromyces lactis*, et *Debaryomyces hansenii*) montre que les sites sont globalement similaires, mais que leur maturation par les facteurs protéiques est régulée par des séquences non conservées situées de part et d'autres. La conséquence est que les facteurs d'une espèce exprimée chez une autre espèce vont produire un profil de polyadénylation très comparable à celui produit par les facteurs sauvages. De plus, l'étude montre que *K. lactis* et *S. cerevisiae* ont naturellement des profils de polyadénylation similaires. Ces deux arguments me conduisent donc à penser qu'une séquence de Cyc1 de *S. cerevisiae* sera correctement liée par les facteurs de *K. lactis*. Un autre point à déterminer vis-à-vis de la séquence à utiliser est sa longueur. Deux raisonnements me semblent possibles à ce point. Je peux prendre le parti de travailler avec une séquence courte, à la manière de Leeper et al. (Leeper, Qu et al. 2010), ce qui limiterait fortement la mobilité amenée par l'ARN en s'assurant que la quasi-totalité de la séquence est engagée dans une interaction (Figure 51). Cependant, comme je l'ai détaillé précédemment, Leeper et al. travaillaient uniquement avec les RRM. Dans mon système, il faut que la séquence soit capable de lier les quatre RRM, le tout en gérant la contrainte spatiale que représente un complexe de la taille du CF IA. C'est pour cette raison que j'ai pris le parti d'utiliser une séquence plus longue, s'étendant du EE au PE. Avec une séquence de cette longueur, le nombre de répétition UA sur l'ARN Gal7 de *S. cerevisiae* double, en comparaison avec la séquence utilisée par Leeper et al., augmentant d'autant les possibilités d'interaction pour Hrp1. Nous utiliserons donc la séquence de Cyc1 de *S. cerevisiae*, plus variée, et qui devrait permettre des interactions plus spécifiques.

A

Séquence de Cyc1sc UUUUAUGUUAUGUUAGUAUUAGAACGUUAUUUAUUAUUCAA
 Séquence de Gal7sc UAUUAUAUAUAUAUAUAUAUAUGACAUCAUUAUCUGUAA
 Séquence de Leeper UAUUAUAUAUAU
 Séquence sélectionnée UUUUAUGUUAUGUUAGUAUUAGAA

B

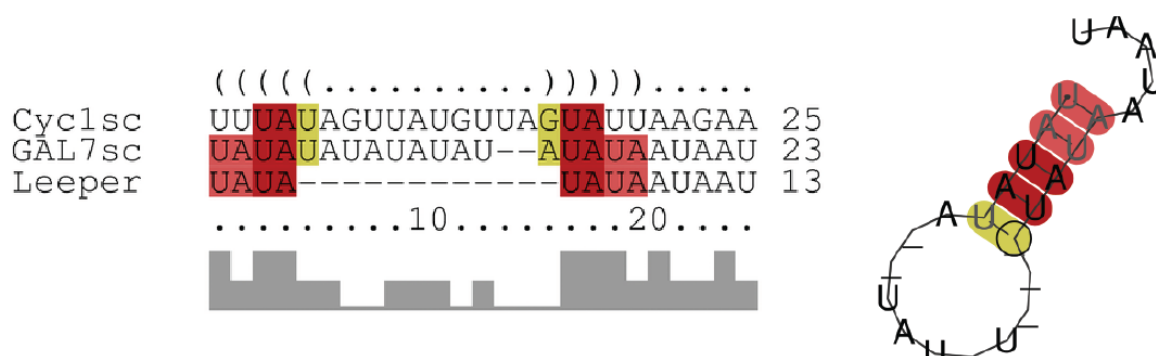


Figure 51 : Comparaison des ARNs communément employés pour les essais fonctionnels. **A.** Séquences des ARNs. L'Efficiency Element, lié par Hrp1, est représenté en rouge. Le Positioning Element, lié par le CF IA, est représenté en bleu. Le site de polyadénylation est représenté en vert. **B.** Prédiction de structure secondaire conservée au sein de l'alignement des séquences (RNAfold)

Une analyse rapide de notre séquence montre qu'elle aurait tendance à former une tige-boucle. L'impact des structures secondaires dans la sélection des sites de polyadénylation est un mécanisme principalement étudié dans le début des années 2000 (Hans and Alwine 2000, Wu and Alwine 2004, Loke, Stahlberg et al. 2005) à propos duquel des exemples sont connus sans qu'une généralisation n'ait été établie. Cette apparente spécificité conduit les auteurs à penser que les structures secondaires n'entreraient pas dans la définition du site de polyadénylation canonique, mais pourraient jouer un rôle important dans les mécanismes de sélection et de polyadénylation alternative. Dans notre cas, les deux régions impliquées dans la formation du double brin sont également supposées interagir avec les protéines. Pour limiter cette compétition, nous dénaturerons l'ARN avant de la mettre en présence des protéines.

3.3.2. Essais de cristallogénèse

Pour effectuer ces essais, les partenaires sont mis en présence à quasi équimolarité (concentrations indiquées dans le Tableau 18) puis incubés une heure à 4°C. Les plaques sont ensuite réalisées à l'aide du robot de cristallisation Mosquito, comme à l'accoutumée.

Construction	Concentration Molaire (g.l)	Conditions testées	T° testées
Cyc1sc RNA Hrp1kl (183-349) CF IAkl His-Pcf11 (322-518)	ARN : $2,07.10^{-5}$ Hrp1 : $2,35.10^{-5}$ (0,45) CF IA : $2,14.10^{-5}$ (6,12)	Cryo PEGs PEGs II Wizard JCSG+ Classics Protein Complex Morpheus MbClass MbClass II PACT Anions Cations Crystal Screen Salt Rx	20°C

Tableau 18 : Résumé des conditions explorées pour la cristallogénèse du CF IA.

Les plaques sont ensuite observées à J+3, J+7 et J+18, sans qu'aucune nucléation n'ait pu être observée jusqu'à présent. Ces essais sont cependant à poursuivre en essayant notamment d'augmenter la quantité de matériel.

3.4. Etude des motifs en « doigts de zinc » de Pcf11p

3.4.1. Analyse in silico

La première et unique mention de motifs liant le zinc sur Pcf11 se trouve dans un article de Sadowski et al.(Sadowski, Dichtl et al. 2003), et il ne s'agit que d'une représentation des domaines hypothétiques sur Pcf11p. Il n'en n'est pas fait mention dans le texte, et aucune étude ne s'y est directement intéressée.

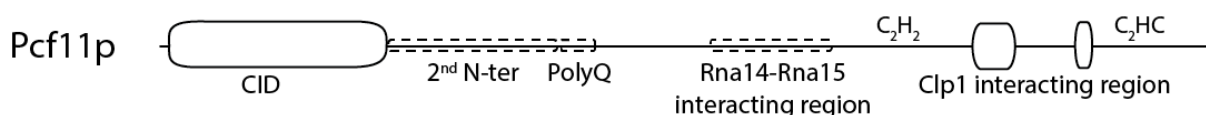


Figure 52 : Les doigts de zinc putatifs dans le contexte des domaines connus de Pcf11p.

J'ai donc commencé par faire un alignement des séquences protéiques des différents homologues de Pcf11. J'ai ainsi récupéré 6 homologues du règne Champignons, 9 de l'embranchement *Chordata* et 2 de l'embranchement *Arthropodia*. Les séquences ont ensuite été alignées à l'aide de ClustalΩ (Goujon, McWilliam et al. 2010, Sievers, Wilm et al. 2011,

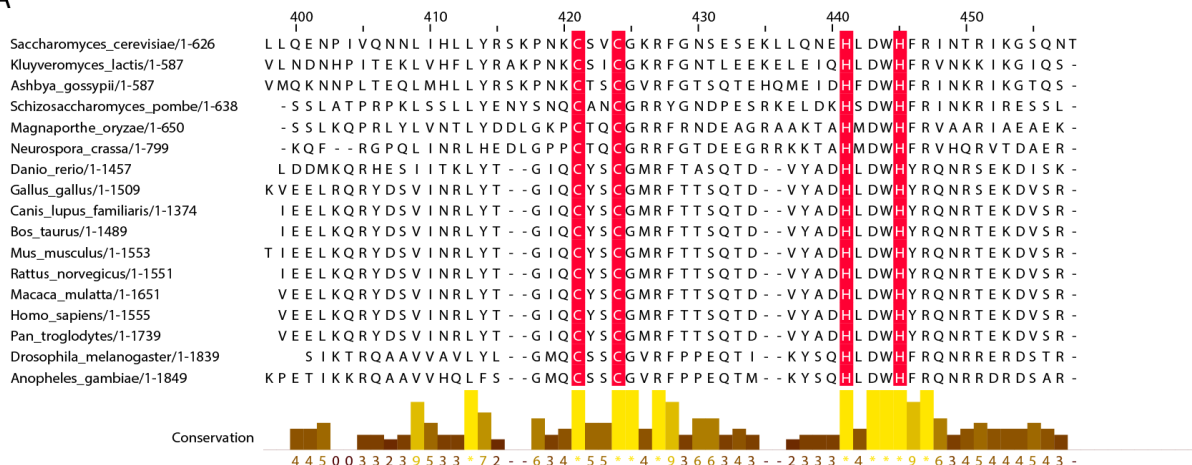
McWilliam, Li et al. 2013), et analysées sur JalView (Waterhouse, Procter et al. 2009). D'une manière générale, il faut remarquer que les séquences du règne Champignon sont particulièrement courtes par rapport à leurs homologues, avec une moyenne autour de 650 résidus, contre respectivement 1550 et 1850 pour les *Chordata* et les *Arthropodia*. Trois domaines sont conservés au sein des trois groupes (Figure 53). Le domaine N-terminal, déjà décrit à de nombreuses reprises, interagit avec le CTD de l'ARN polymérase II (Noble, Hollingworth et al. 2005, Schrieck, Easter et al. 2014). Les deux domaines suivants, interagissant respectivement avec Rna14-Rna15 et avec Clp1, sont moins bien connus. Dans ces domaines, on peut remarquer que seuls quelques acides aminés sont parfaitement conservés. La Figure 54 représente un agrandissement des régions concernées, et nous pouvons remarquer que les résidus conservés évoquent clairement un motif en doigt de zinc relativement canonique.

Le premier motif, de type Cys₂His₂, semble avoir toutes les caractéristiques théoriques de la classe de doigts de zinc la mieux caractérisée. Simplement repliés en $\beta\beta\alpha$, ils sont canoniquement décrits par la séquence X₂-Cys-X_{2,4}-Cys-X₁₂-His-X_{3,4,5}-His (Pabo, Peisach et al. 2001) et interagissent avec un seul atome de zinc. Le second motif est caractérisé par les résidus Cys₂HisCys, qui n'est pas un motif particulièrement représenté.



Figure 53 : Vue générale de l'alignement des homologues de Pc11. Les alignements sont réalisés avec ClustalΩ.

A



B

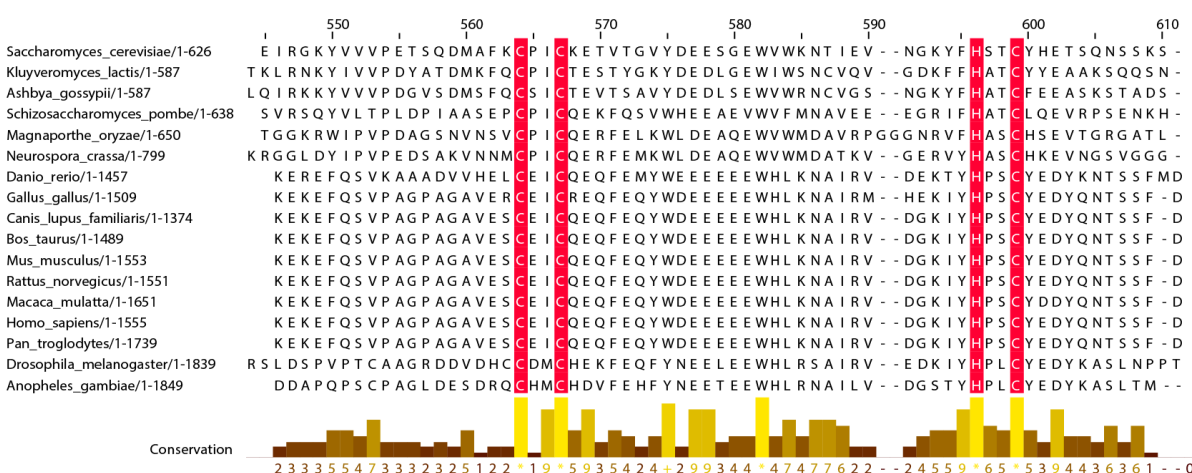


Figure 54 : Alignement des homologues de Pf11. Les alignements sont réalisés avec ClustalΩ, seule la numérotation de *Saccharomyces cerevisiae* est indiquée. La conservation est représentée par une échelle de 0 à 10 (*) A. Agrandissement dans la région 398-466. On peut observer que les résidus C421, C424, H441 et H445 identifiés en rouge sont particulièrement conservés. B. Agrandissement dans la région 544-611. Il faut ici remarquer la conservation des résidus C564, C567, H596 et H599.

Cette piste semblant intéressante à explorer, j'ai donc décidé de poursuivre.

3.4.2. Absorption atomique et liaison de zinc

Nous suspectons maintenant fortement l'existence de doigts de zinc sur Pf11p, mais comment pourrions-nous nous assurer qu'ils existent réellement ? Nous avons décidé de recourir à la spectrométrie d'absorption atomique au seuil du zinc, et je remercie le Dr Yves Méchulam de l'Ecole Polytechnique d'avoir gracieusement effectué ces mesures. Très sensible, cette technique est extrêmement rapide à mettre en œuvre, quantitative et peut être effectuée sur l'échantillon en solution. La seule limite est la quantité de matériel à fournir, qui est relativement importante. A cause de cela, nous avons choisi de travailler avec trois

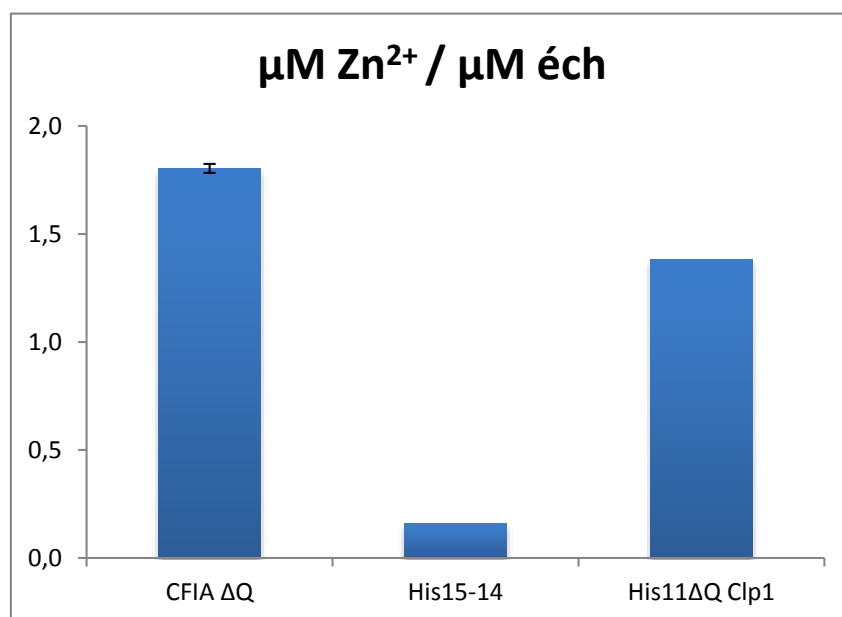
échantillons que nous connaissions bien et que nous sommes en mesure de produire en quantité et en qualité suffisante. Nous avons donc analysé un échantillon de His-Rna15p-Rna14p, un échantillon de His-Pcf11pΔQ-Clp1p et deux échantillons de CF IA ΔQ.

En pratique, une mesure est effectuée sur les échantillons et sur le tampon seul, afin soustraire sa contribution. La valeur d'absorbance est alors reportée sur une gamme d'absorbance de ZnSO₄ afin d'obtenir la concentration de Zinc présente dans notre échantillon.

NOM	PROTEINES	CONCENTRATIO N PROTEINE (μM)	CONCENTRATIO N ZN (μM)	μM ZN / μM ECH
HIS-Rna15- Rna14	HisRna15p + Rna14p	3,14	0,5	0,16
CFIA ΔQ	His Pcf11pΔQ + Rna15p + Rna14p + Clp1p	6,6	12	1,82
CFIA ΔQ	His Pcf11pΔQ + Rna15p + Rna14p + Clp1p	6,15	11	1,79
HISPCF11Δ Q CLP1	His Pcf11p deltaQ20 + Clp1p	17	23,5	1,38

Tableau 19 : Concentration en Zinc mesurée dans les échantillons. La concentration en protéine est estimée à partir de mesure de l'*Abs_{280nm}* mesurée sur Nanodrop. La concentration en zinc est calculée à partir des mesures d'*Abs_{213,9nm}*.

Dans le Tableau 19, nous pouvons retrouver les valeurs en concentration de protéine et en concentration de zinc calculées à partir de mesures d'absorbance. Qualitativement parlant, on doit remarquer que du zinc est retrouvé dans les échantillons HisPcf11pΔQ-Clp1p et CF IA ΔQ, alors que la mesure effectuée sur His-Rna15p-Rna14p n'est pas significativement différente du seuil de détection. Nous savons, grâce à la structure de Clp1p (Noble, Beuth et al. 2007, Dupin and Fribourg 2014), qu'il n'y a pas de zinc dans Clp1p, ce qui signifie de le zinc détecté dans CF IA est apporté en intégralité par Pcf11p. Cependant, comme nous ne pouvons pas exprimer Pcf11p seul, nous ne pouvons pas nous assurer que l'interaction entre Pcf11p et Clp1p n'est pas responsable de la fixation du Zinc, chacun des deux partenaires pouvant apporter une partie du site de fixation. Nous pouvons néanmoins considérer que les motifs formés par les résidus conservés s'organisent très probablement en domaine liant le zinc.



Graphique 1 : Proportion de Zinc retrouvé dans les échantillons de protéines.

Nous avons ensuite rapporté la quantité de zinc mesurée dans l'échantillon à la quantité de protéine estimée par la mesure d'absorbance dans la dernière colonne du Tableau 19 et dans le Graphique 1. Nous garderons à l'esprit qu'il n'est pas possible de mesurer précisément la concentration en protéine par cette technique, et qu'il ne s'agit donc là que d'une estimation. Nous voyons que nous avons environ 1,8μM de zinc pour 1μM de CF IA, ce qui semble aller dans le sens des deux doigts de zinc présents dans le complexe. S'il est indéniable qu'il y a du zinc dans HisPcf11pΔQ-Clp1p, le rapporter à la quantité de protéine nous donne un rapport de 1,38 qui est plus difficile à interpréter. En effet, Pcf11p étant relativement peu stable, cela signifierai que notre mesure surestimerai la concentration réelle, faussant le rapport. Cela laisserait penser que la liaison au zinc est moins bonne hors du contexte de CF IA complet, mais nous n'avons pas d'éléments pour expliquer ou confirmer ceci.

3.4.1. Analyse en MS

Nous savons que Pcf11p est chargé en zinc, et notre hypothèse est que les motifs formés par les résidus conservés précédemment identifiés sont responsables de cette interaction via une organisation en doigt de zinc. Pour confirmer cela, nous avons cloné et exprimé des fragments de Pcf11p contenant ces motifs afin de les analyser en spectrométrie de masse native pour caractériser leur contenu en zinc et la stœchiométrie précise.

A cause de soucis d'ordre techniques, les premières analyses effectuées sur le premier doigt de zinc ne sont pas exploitables, et ne pourrons donc pas être discuté dans ce documents.

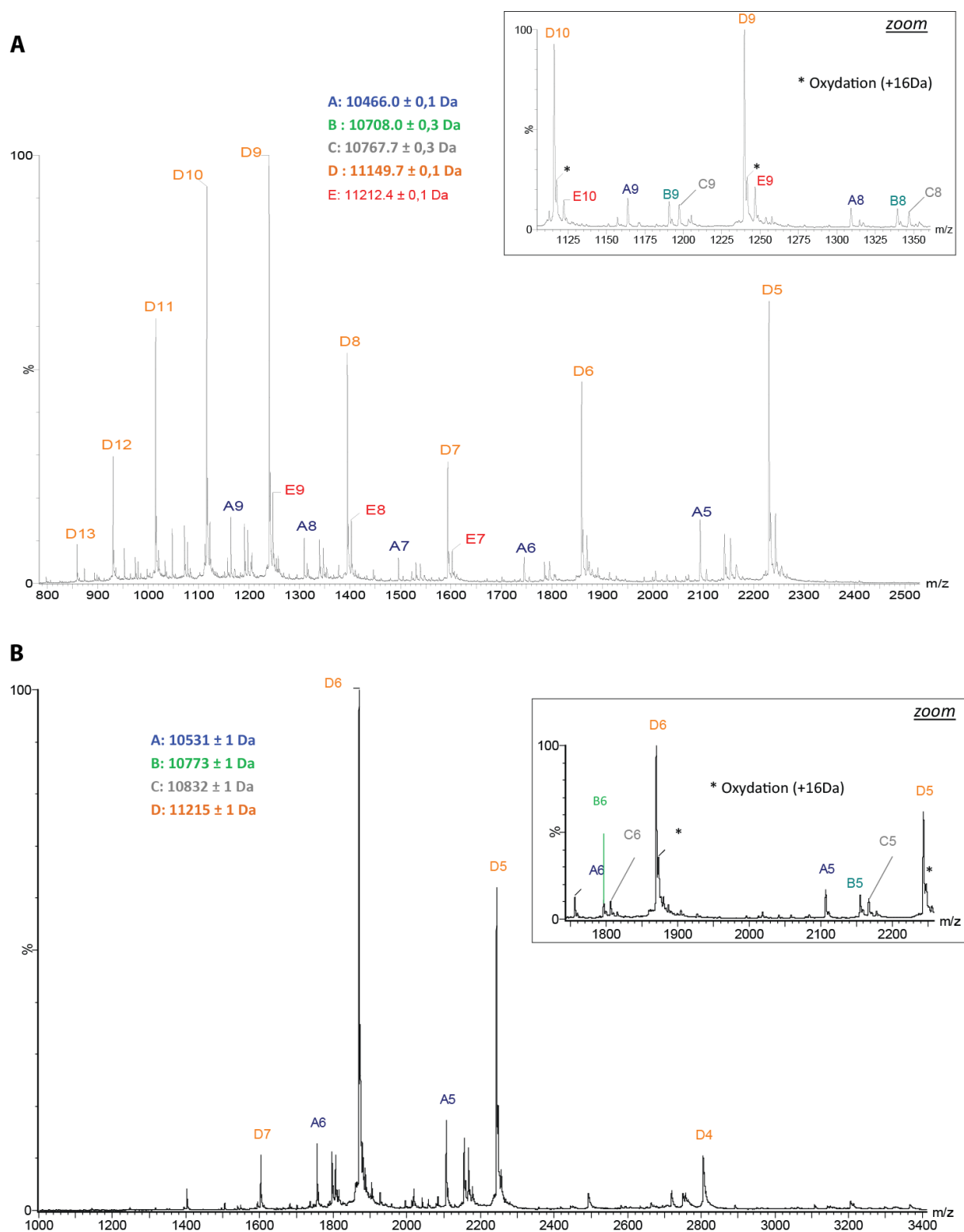


Figure 55 : Analyse du fragment Pcf11p (530-626) par spectrométrie de masse en conditions dénaturantes (A) ou natives (B)

La Figure 55 présente les résultats obtenus sur le fragment Pcf11p (530-626). Ce fragment, contenant le deuxième doigt de zinc potentiel, est exprimé en fusion avec un tag His6-GST-TEV et purifié par une étape d'affinité contre la GST. Le tag est alors clivé et le fragment est ségrégué par exclusion de taille, générant la séquence suivante : (les résidus en gras sont apportés par l'étiquette)

GHMGSVDNTLGSDRSNELEIRGKYVVVPETSQDMAFK**CPICK**ETVTGVYDEESGEVWVKNTIEVNGKY
FHSTCYHETSQNSSKSNSGKVGGLDDLKLVTK*

Le panneau A de la Figure 55 représente l'analyse de l'échantillon en conditions dénaturantes et on voit que 5 espèces distinctes, reprises dans le Tableau 20, sont détectées.

Espèces	Masse mesurée (Da)	Protéine assignée	Masse théorique (Da)	Différence de masse (Da)	Explications
A	10 466	Pcf11p (530-626)	11 150,3	- 684,3	Dégradation
B	10 708	Pcf11p (530-626)	11 150,3	- 442,3	Dégradation LVTK C-terminaux
C	10 767,7	Pcf11p (530-626)	11 150,3	- 382,6	B + 59,7 Da = B + Zn ²⁺ ou Dégradation GHMG N-terminaux
D	11 149,7	Pcf11p (530-626)	11 150,3	-0,6	
E	11 212	Pcf11p (530-626)	11 150,3	+ 61,7	D + 62,3 Da = D + Zn ²⁺

Tableau 20 : *Espèces détectée par l'analyse de Pcf11p (530-626) en conditions dénaturantes*

L'espèce D, largement majoritaire, est la protéine non dégradée. Les espèces A, B et C peuvent être assimilées à des produits de dégradation ou, pour C, à une forme dégradée liée à un atome de Zinc. La forme E pourrait elle aussi correspondre à une espèce liée à un ion Zn²⁺, mais il s'agirait cette fois de la protéine entière.

Si nous nous intéressons maintenant à l'analyse effectuée en conditions natives, présentée dans le panneau B de la Figure 55, nous voyons que seulement quatre espèces sont détectées. Ces espèces sont résumées dans le Tableau 21.

Espèces	Masse mesurée (Da)	Protéine assignée	Masse théorique (Da)	Différence de masse (Da)	Explications
A	10 531	Pcf11p (530-626)	11 150,3	- 619,3	$A^{\text{dénat}} + 65 \text{ Da} = A^{\text{dénat}} + \text{Zn}^{2+}$
B	10 773	Pcf11p (530-626)	11 150,3	- 377,3	$B^{\text{dénat}} + 65 \text{ Da} = B^{\text{dénat}} + \text{Zn}^{2+}$
C	10 832	Pcf11p (530-626)	11 150,3	- 318,3	$C^{\text{dénat}} + 64,3 = C^{\text{dénat}} + \text{Zn}^{2+}$
D	11 215	Pcf11p (530-626)	11 150,3	+ 64,7	$D^{\text{dénat}} + 65,3 = D^{\text{dénat}} + \text{Zn}^{2+}$ $= E^{\text{dénat}}$

Tableau 21 : Espèces détectée par l'analyse de Pcf11p (530-626) en conditions natives

En comparant les espèces visualisées en conditions natives et dénaturantes, nous pouvons immédiatement remarquer que toutes les espèces natives sont liées à une espèce dénaturée par une variation de masse très proche de 65 Da. Cela nous permet de conclure qu'il y a bien un atome de zinc fixé sur ce fragment de Pcf11p, traduisant son organisation en doigt de zinc.

3.4.2. Mutants Test in vivo & in vitro

Comme nous l'avons expliqué précédemment, aucune étude n'a jamais été menée sur l'existence de motifs liant le zinc, et donc *a fortiori* sur leur fonction biologique. Nous savons uniquement qu'ils encadrent approximativement la région d'interaction avec Clp1p, sans pour autant y participer. Pour explorer leur fonction, j'ai dessiné des mutants perturbants les motifs en modifiant les deux premières cystéines en sérines par mutagenèse dirigée.



Figure 56 : Représentation des mutations intégrées dans la séquence de Pcf11p. Les résidus mutés sont représentés en rouge, les résidus inchangés en bleu.

Nous appellerons Pcf11p Zn1 le mutant portant les mutations C421S et C424S interrompant le premier motif. Le mutant Pcf11p Zn2 perturbe le second motif grâce aux mutations C564S et C567S. Un double mutant Pcf11p Zn1-Zn2 portant les quatre mutations est également cloné afin de perturber les deux motifs.

Une fois ces mutants en ma possession, je me suis attaché à reconstituer CF IA *in vitro* par surexpression des protéines recombinantes. Malgré plusieurs tentatives, je ne suis pas parvenu à obtenir des co-purifications réellement exploitables, principalement à cause de la dégradation subie par Pcf11p, caractéristique de la surexpression de la protéine entière. Ceci m'empêche également de tester le contenu en zinc sur des séquences mutées afin de m'assurer qu'elles ne sont plus compétente pour fixer le zinc. J'ai donc cloné les mutants dans des plasmides pFL36 permettant l'expression chez la levure en collaboration avec Julia Guegueniat. Après génération de souches où la copie sauvage de Pcf11p est remplacée par le mutant, une purification grâce à l'étiquette TAP fusionnée à la sous-unité Rna15p nous permet de purifier CF IA. Le gel présenté en Figure 58 est une coloration à l'argent des purifications de CF IA à partir d'une souche avec un Pcf11p sauvage, un Pcf11p Zn1 et un Pcf11p Zn2. Et nous pouvons voir que dans tous les cas, nous sommes capables de purifier un CF IA complet, ce qui signifie que les doigts de zinc n'interviennent pas dans la formation de CF IA. Ce résultat était quelque peu anticipé car comme nous l'avons déjà souligné, les motifs se situent à l'extérieur des régions décrites comme impliquées dans les interactions entre Pcf11 et Rna14p-Rna15p de CF IA.

Cependant, ce n'est évidemment pas parce que la disruption des doigts de zinc n'impacte pas la formation de CF IA qu'elle n'a pas d'impact sur la biologie de la levure. Pour explorer cet aspect, nous exploitons le fait que Pcf11p soit un gène essentiel de la levure, ce qui nous permet d'effectuer des tests de croissance comparant la souche sauvage et les souches mutantes (Figure 57). En nous intéressant à la croissance des colonies à 30°C, nous pouvons remarquer qu'il n'y a pas de différence significative entre la souche sauvage et les mutants. Cette absence de phénotype mutant signifie que les protéines sont capables d'assurer leur fonction de manière totale ou partielle, mais cela n'affecte pas la croissance. En observant la croissance de levure à d'autres températures, nous voyons que les mutants montrent une thermosensibilité légère à 37°C pour le Zn1 mais plus marquée pour le Zn2. Nous pouvons interpréter cela comme une diminution des capacités d'adaptation de la levure à son environnement.

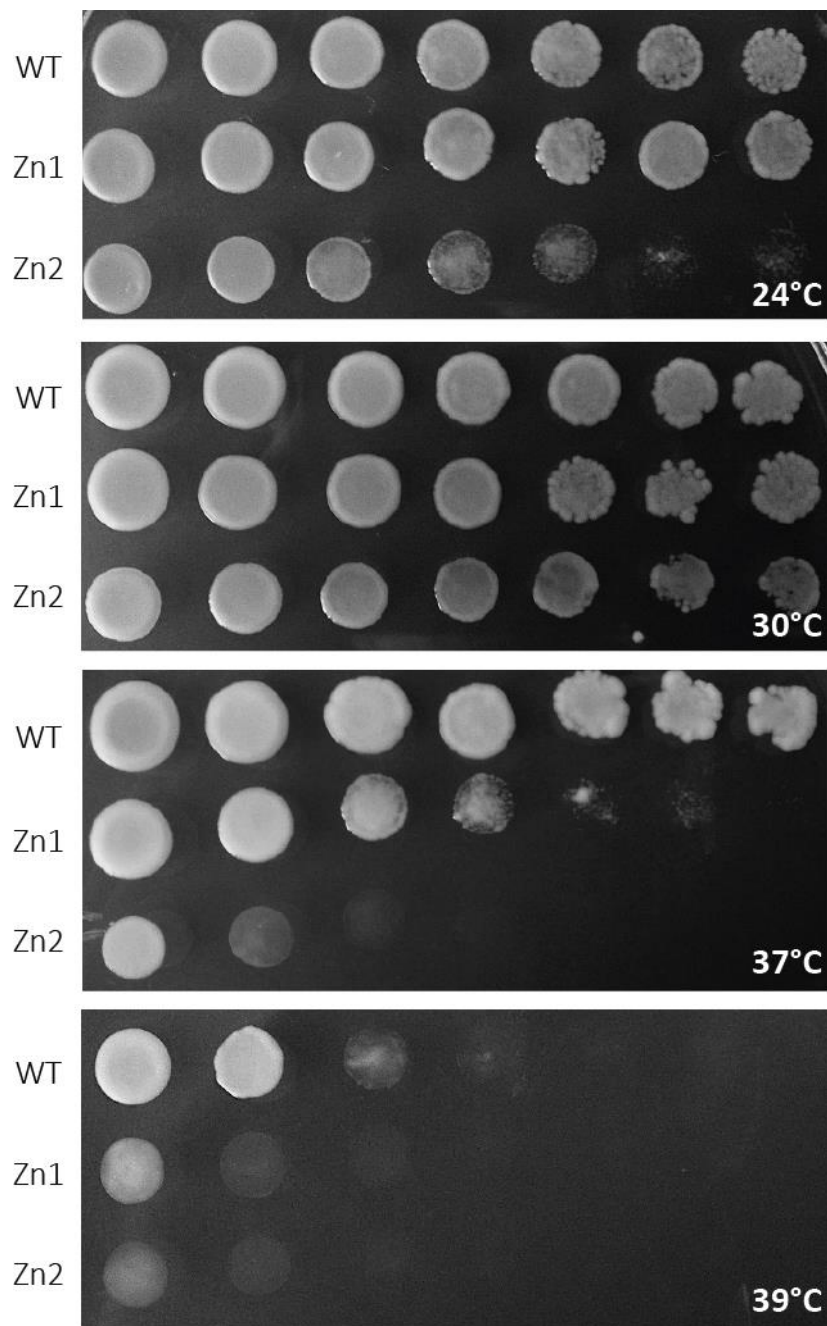
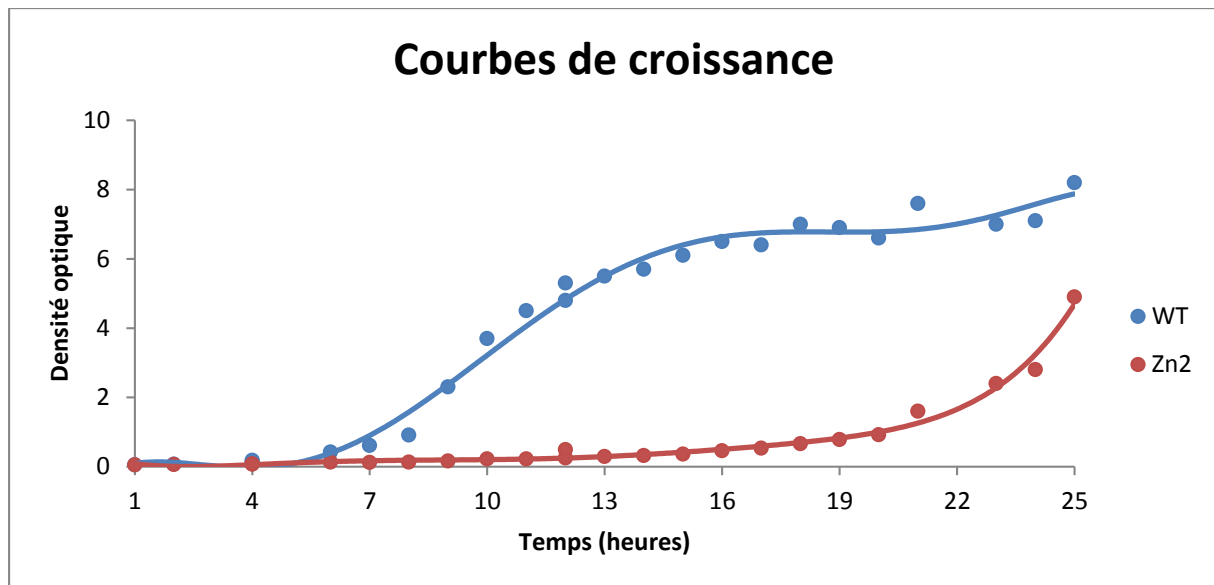


Figure 57 : Test de croissance en goutte pour des souches de *S.c.* portant différentes copies de *Pcf11p*. 30°C est la température de croissance optimale. 24°C, 37°C et 39°C sont des températures plus stringentes pour la levure. La croissance est effectuée sur milieu YPD(A) pendant 24 à 48h. Réalisé par Julia Guegueniat, U869.

Le mutant Zn2 (C564S/C567S) semble tout particulièrement affecté. Pour mieux le caractériser, nous avons suivi la croissance de cette souche en culture liquide à 30°C (Graphique 2).



Graphique 2 : Courbes de croissance enregistrées pour la souche sauvage (bleue) et la souche exprimant l'allèle mutant *Pcf11p Zn2* (rouge). La culture est effectuée sous agitation, à 30°C, en milieu YPD(A). La densité optique est mesurée à 600nm. Réalisé par Julia Guegueniat, U869.

Grâce à cette cinétique de croissance, nous voyons très bien le retard de croissance subit par le mutant *Zn2*, même à 30°C. La souche sauvage atteint une densité optique de 4 environ 10 heures après la mise en culture, alors que le mutant a besoin de 24h pour arriver à cette population. La double mutation C564S/C567S affecte la biologie de la levure, diminuant sensiblement son aptitude à l'adaptation à l'environnement. Nous ne pouvons pas savoir d'où vient cette perte d'efficacité cependant. La perte du doigt de zinc entraîne-t-elle la perturbation d'interactions entre *Pcf11p* et l'un de ses nombreux partenaires ? Ou bien est-ce une déstructuration du domaine qui est intrinsèquement responsable de la perturbation des domaines voisins ?

Pour amener des éléments de réponse, nous allons purifier et tester l'activité de complexes CF IA portant ces mutations pour le clivage et la polyadénylation en collaboration avec Julia Guégueniat. Pour l'heure, seule la purification a pu être effectuée (Figure 58). Si le facteur sauvage et le mutant *Zn1* ont des profils similaires, nous pouvons remarquer que les signaux sont plus discrets chez le mutants *Zn2*. Toutes les sous-unités sont visibles néanmoins, indiquant que l'association du CF IA n'est pas défectueuse, mais elle semble perturbée, ce qui expliquerait le défaut de croissance.

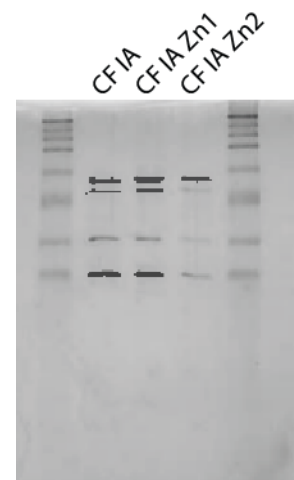


Figure 58 : Purification des facteurs de levure.

3.4.3. Essais de cristallisation

Dès le moment où nous avons confirmé la présence de zinc au sein de CF IA, nous avons cloné et exprimé différentes constructions correspondant aux deux doigts de zinc afin de les caractériser en spectroscopie de masse, mais également afin de tenter de résoudre leur structure. Trois constructions ont ainsi été développées (388-459 ; 403-459 et 530-626) mais seulement deux ont pour l'instant pu être testées pour la cristallogénèse.

Constructions testées	Conditions testées	Concentrations testées (mg/ml)	T° testées	Hits
Pcf11p (388-459)	Classics	7,3	20°C	
Pcf11p (530-626)	Anions Cristal Screen JCSG+ PEG I PEG II Mb Class Wizard Index Classic	9,4 10,2 15,2 18,5 25,5	20°C	Oui

Tableau 22 : Essais de cristallogénèse effectués sur les doigts de Zinc.

Des cristaux ont été obtenus dans certaines conditions, des optimisations sont maintenant en cours afin d'essayer d'acquérir des jeux de données sur ces domaines.

3.5. Mise en évidence d'une région d'interaction étendue et conservée entre Cpl1p et Pcf11p

3.5.1. La structure de Noble et les questions en suspens

Le rôle de Clp1p dans la machinerie de la maturation 3' des ARNm est relativement peu connu. Cette sous-unité interagit avec Pcf11p au sein du CF IA, mais également avec des protéines du CPF. La structure de Clp1p en complexe avec un fragment de Pcf11p a été résolue (Noble, Beuth et al. 2007), mettant également en évidence la présence d'ATP au sein de la protéine. Cette interaction avec le nucléotide est conservée parmi les espèces bien que l'activité d'hydrolyse ne soit détectée que chez les eucaryotes supérieurs et les archées, et liée non pas à la maturation des ARNm mais à celles des ARNt. Bien qu'elle démontre la

présence d'ATP, la résolution de la structure n'a pas permis d'expliquer le phénotype de certains mutants de Clp1p perturbés dans la liaison à l'ATP et/ou l'interaction avec Pcf11p (Ghazy, Gordon et al. 2011, Haddad, Maurice et al. 2011, Holbein, Scola et al. 2011).

3.5.2. Les essais de cristallographie

Le tableau suivant présente uniquement les essais que j'ai effectué moi-même, et non la totalité des essais réalisés sur ce projet.

Projet	Constructions testées	Conditions testées	Concentrations testées (mg/ml)	T° testées	Hits
Interaction Clp1-Pcf11 (cf 3.5)	His-Pcf11p (454-563) Clp1p	PACT Screen PEG 3000 Screen PEG 3350 Screen PEG 4000 Screen PEG 5000 Screen PEG 6000 Screen Glycérol Screen Ethylène Glycol Screen pH Tris	5	20°C	oui
Explication moléculaire des mutants de Clp1 (cf 3.5)	His-Pcf11p (454-563) Clp1p-6	Additifs Screen pH Acide Citrique Screen pH Tris Screen pH BTP Screen PEG 3000 Screen PEG 3350 Screen PEG 4000 Screen PEG 5000 Screen PEG 6000 Screen PEG 8000	5 10	20°C	

Tableau 23 : Récapitulatif des essais de cristallogénèse pour la caractérisation des mutants de Clp1p et de l'interaction entre Clp1p et Pcf11p.

3.5.3. Article (PMID : 24508575)



Contents lists available at ScienceDirect

Biochimie

journal homepage: www.elsevier.com/locate/biochi

Research paper

Structural basis for ATP loss by Clp1p in a G135R mutant protein

Adrien F. Dupin^{a,b}, Sébastien Fribourg^{a,b,*}^a Univ. Bordeaux, IECB, F-33607 Pessac, France^b INSERM, U869, F-33077 Pessac, France

ARTICLE INFO

Article history:

Received 8 November 2013

Accepted 17 January 2014

Available online 5 February 2014

Keywords:

CF IA

3'-end

Pre-mRNA

ABSTRACT

Pcf11p and Clp1p form a heterodimer and are subunits of the Cleavage Factor IA (CF IA), a complex that is involved in the maturation of the 3'-end of mRNAs in *Saccharomyces cerevisiae*. The role of Clp1p protein in polyadenylation remains elusive, as does the need for ATP binding by Clp1p. In order to obtain structural details at atomic resolution of point mutants of Clp1p, we solved the crystal structure of Clp1-1p (G135R) point mutant complexed with Pcf11p (454–563) domain. The Clp1-1p–Pcf11p structure provides the atomic details for ATP loss while the point mutation preserves intact the Pcf11p interaction surface of Clp1p. This provides a rationale for the absence of phenotype in the yeast *clp1-1* strain. Additionally, the structure allows for the description of an extended binding interface of Pcf11p with Clp1p which is likely to be *S. cerevisiae* specific.

© 2014 Elsevier Masson SAS. All rights reserved.

1. Introduction

Polyadenylation is a co-transcriptional event that occurs at the 3'-end of nearly all eukaryotic mRNA precursors with the exception of histone mRNAs [1,2]. This essential maturation process starts with cleavage of the elongating pre-mRNA in the 3'-untranslated region (3'-UTR) followed by the addition of a poly(A) tract at the 3'-hydroxyl end of the upstream cleavage product, the 3' fragment being degraded by Rat1p/hXrn2 [3,4]. Pre-mRNA processing and transcription termination rely on the recognition of *cis*-elements by the polyadenylation machinery [1,5,6].

In yeast, the polyadenylation machinery comprises two large multi-protein complexes: the Cleavage/polyadenylation Factor IA (CF IA) [7] and the Cleavage and Polyadenylation Factor (CPF). In addition other factors are required to complete the process: the RNA-binding protein Hrp1p/Nab4p [8,9] and the poly(A)-binding proteins Pab1p and Nab2p, the latter being involved in the control of poly(A) tail synthesis and subsequent export of the mature mRNA [7,10–13]. The four CF IA subunits are required to fulfil CF IA function in 3'-end mRNA processing [14]. In addition to its role in the 3'-end mRNA maturation process, CF IA is also involved in transcription termination [15–17] and reinitiation [18]. CF IA consists of Rna14p, Rna15p, Pcf11p and Clp1p. Rna14p is a TPR-containing protein that acts as a scaffold protein organizing the overall architecture of CF IA [14]. It binds Rna15p through its C-

terminal monkey tail region [19]. Through its homodimerization, Rna14p most probably attracts two copies of Rna15p [20–22]. Rna15p is a modular protein containing at its N-terminus a classical RBD [23] displaying high affinity for U-rich sequences but specifically binding A-rich sequences upstream to the poly-A site, in the presence of Hrp1p [23–26]. The middle region of Rna15p, also called Hinge domain, binds tightly to Rna14p [19] through a co-folding process and the C-terminus of Rna15p binds to Pcf11p [27]. Pcf11p is required for both pre-mRNA 3'-end processing and transcription termination. The CTD-interacting domain (CID) of Pcf11p is dispensable for 3'-end processing *in vitro* but it is essential to avoid transcriptional read-through *in vivo* [17,28]. Indeed, the Pcf11p CID interacts specifically with the Ser2-phosphorylated form of the carboxyl-terminal domain (CTD) of the largest subunit of RNA polymerase II (Pol II) [17,29], while the central and C-terminal regions of Pcf11p contact Rna14p/Rna15p and Clp1p, respectively [28,30]. Finally, Clp1p is an ATP binding protein bearing classical Walker motif [30]. The association of Clp1p to the remainder of CF IA subunits is mainly restricted to a tight interaction with Pcf11p [30]. The heterotetramer of Rna14p–Rna15p associates with only one copy of the Pcf11p–Clp1p heterodimer. The molecular basis for such asymmetry has not been addressed so far.

Even though Clp1p binds ATP, the precise role of this protein in 3'-end processing remains elusive. In higher eukaryotes and archaea, Clp1 proteins are kinases that phosphorylate 5' hydroxyl ends of RNA [31,32]. In humans, the loss of its kinase activity is connected to progressive motor-neuron loss and connected to tRNA processing rather than 3'-end mRNA maturation alteration [33]. However, this kinase activity is restricted to higher eukaryotes and archaea, while no similar activity could be detected for yeast Clp1p

* Corresponding author. INSERM, U869, F-33077 Pessac, France. Tel.: +33 5 4000 3063.

E-mail address: sebastien.fribourg@inserm.fr (S. Fribourg).

[30–32]. It has been shown that addition of ATP stimulates 3'-end mRNA processing and that the cleavage reaction itself can proceed in the presence of CTP and 3' dATP [9,34,35].

On the basis of recent studies that report extensive and independent mutational analysis of Clp1p [36–38], we have undertaken a structural characterization of some of these point mutations. From the different experiments (Clp1-1p (G135R), Clp1-5p (G135R/G335S), Clp1-6p (G335S), Clp1-19p (L341S), Clp1-16p (G135R/L341S)) altering either ATP-binding site and/or the Pcf11p binding site [37], we could only determine the structure of Clp1-1p (G135R) point mutant complexed with Pcf11p (454–563), while the other samples led to non-diffracting crystals (Clp1-19p) or unstable complexes (Clp1-6p, Clp1-5p and Clp1-16p). The structure provides the atomic details for ATP-loss in Clp1-1p. No ATP is seen in the remodelled binding pocket while Pcf11p interaction is preserved. We also describe an extended binding interaction of Pcf11p with Clp1p. Due to the poor quality of the electron density map in this area the exact nature of the sequence is difficult to assign. A likely assignment is proposed.

2. Material and methods

2.1. Protein preparation and crystallization

The yeast DNA sequence encoding Pcf11p (454–563) protein fragment and full-length Clp1p were amplified by PCR and inserted into a pET-derived (Novagen) plasmid [39] to produce an N-terminal 6His-tag fused Pcf11 protein fragment and a native Clp1p. Point mutants of Clp1p were generated by PCR and further verified by sequencing. Clp1p mutant proteins and Pcf11p were co-expressed in *Escherichia coli* BL21(DE3) and were further purified to homogeneity by an initial cobalt-affinity chromatography step against the His-tag performed in 25 mM Tris–HCl pH 7.5, 150 mM NaCl (Buffer A) and eluted with Buffer A supplemented with 250 mM imidazole. A Hi-Q Sepharose chromatography column was further performed with Buffer A as loading buffer. Elution was performed by a gradient from 150 mM to 1000 mM NaCl in 25 mM Tris–HCl pH 7.5. The His-tag was removed by an overnight incubation with 1/100 (w/w) TEV protease in a Buffer A supplemented with 0.5 mM EDTA and 1 mM DTT. A final size exclusion chromatography HR26/60 75 µg (GE Healthcare) in 25 mM Tris–HCl pH 7.5, 150 mM NaCl was applied to exchange the buffer and remove excess Pcf11p (454–563) and cleaved His-tag. The same procedure was applied for Clp1p point mutants G135R (Clp1-1p), G335S (Clp1-6p), G135R/G335S (Clp1-5p), G135R/L341S (Clp1-16p), L341S (Clp1-19p).

When possible, the protein samples were concentrated up to 10 mg/ml and crystallized at 20 °C by the hanging drop method with a ratio of 1:1 sample protein versus well with 0.5 µL drops. The best Clp1-1p–Pcf11p (454–563) protein crystals were initially obtained after one day from a solution containing 12% PEG 4000 (w/v), 0.1 M HEPES pH 7.5 or Tris–HCl pH 8.5, 0.1 M–0.2 M MgCl₂.

Clp1-1p–Pcf11p (454–563) complex crystals were cryo-protected in a solution containing 0.1 M Tris–HCl pH 8.5, 25% PEG3350 (w/v), 5% PEG 400 (w/v) and flash-cooled in liquid nitrogen.

2.2. Structure determination

Crystals belonged to the space group $P2_12_12_1$ with two molecules of heterodimer per asymmetric unit (R.m.s.d. of 1.0 Å on α atoms over 428 residues). Datasets were reduced using XDS [40]. We carried out an initial rigid body refinement using Phenix and the Clp1p coordinates from 2NPI PDB prior to further cycles of refinement with Buster and manual model building [41–43].

Statistics of data collection, processing and refinement are summarized in Supplemental Table 1.

3. Results

3.1. Structure of a mutant Clp1p G135R–Pcf11p complex

The wild type and the G135R Clp1p–Pcf11 protein complexes (residues 454–563) yielded well-diffracting crystals and X-ray diffraction data were collected up to 2.77 Å and 2.70 Å respectively (Supplemental Table 1). The structure were obtained using the 2NPI PDB as a starting model and refined to an Rfree/Rfactor of 17.86%/21.30% and 19.06%/22.30%. Clp1p adopts the structure, as described by Noble and co-workers, with a N-terminal domain formed by residues 1–101 (Fig. 1, NtD in red), a C-terminal domain composed of residues 342–445 (CtD, in yellow) arranged around a middle larger domain (Middle, in green) [30]. The first 18 N-terminal residues are also missing in the structure. Overall our Clp1p–Pcf11p (454–563) structure superimposes well on the 2NPI structure with an R.m.s.d. of 0.47 Å on α atoms over 428 residues (Fig. S1). Upon structure refinement a non-attributed extra density appeared located around the C-terminal domain of Clp1p (Figs. 1 and 2). The overall quality of the data and the limited resolution in this area of the electron density map made difficult the unambiguous assignment of the sequence. From an initial OMIT map (Fig. 2a) a sequence of 5 consecutive residues could be built as a α chain. Upon model refinement a total of 9 residues was built and attributed to residues 460–469 of Pcf11p (Fig. 2b). The rationale for this assignment sits in an overall better agreement with X-ray data for this 9-residue stretch in comparison to other trials on the one hand and in the proximity with residue 475 of Pcf11p on the other hand. These residues are probably providing a mild contribution to the interaction with Clp1p, as their deletion does not alter Clp1p pull-down in comparison to wild-type Pcf11p (454–563) fragment (Fig. S2a). This observation is in agreement with the poor residues conservation of this stretch (Fig. S2b). Altogether, the modelled Pcf11p fragment encompasses residues 460–469 and 475–499.

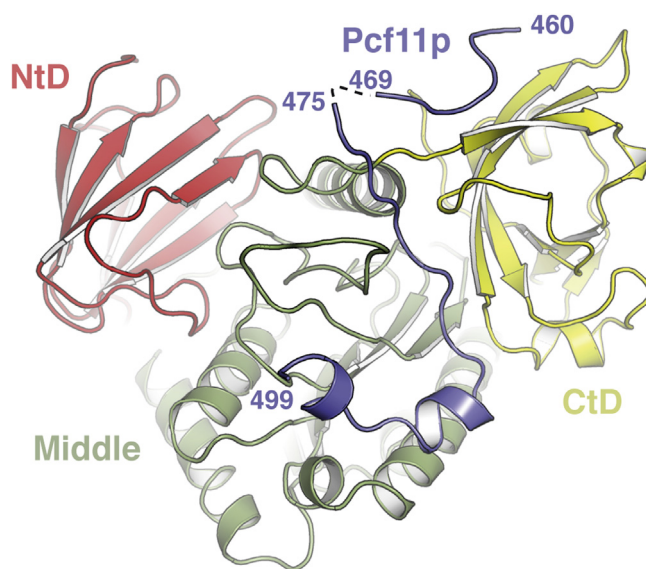


Fig. 1. Structure of Clp1-1p–Pcf11p (454–563). The N-terminal domain (NtD, pink), Middle domain (green) and C-terminal domain (CtD, yellow) are shown as described in Noble et al. [30]. Pcf11p is depicted in blue and the protein boundaries are indicated.

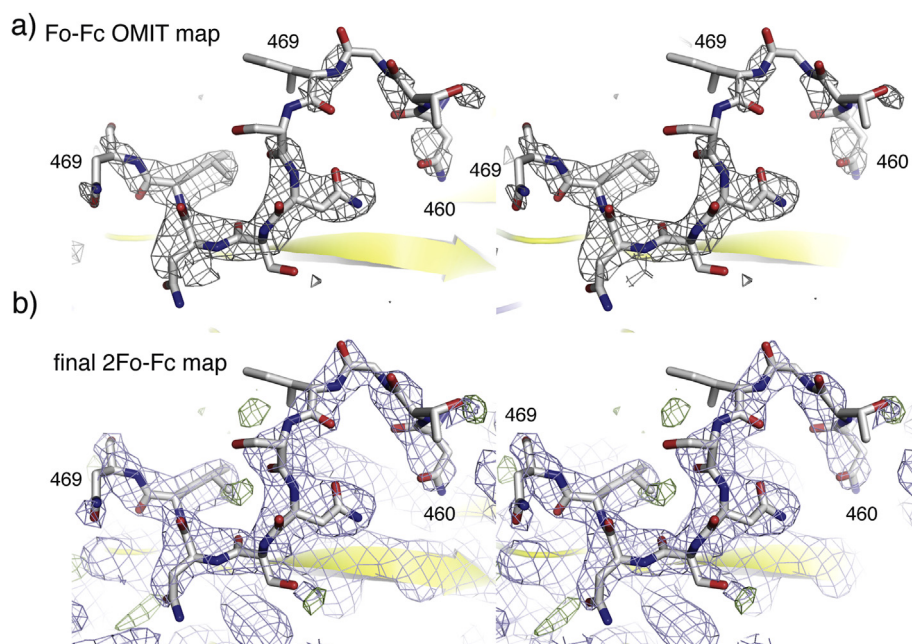


Fig. 2. Extended interaction surface between Pcf11p and Clp1p. a) Stereo diagram displaying the OMIT Fo-Fc (dark, contoured at 3σ) electron density map around residues attributed to Pcf11p (460–469). The final model (omitted in the calculation) is placed into the density. b) Stereo diagram displaying the final 2Fo-Fc (blue, contoured at 1σ) and Fo-Fc (green, at 3σ) electron density maps around residues attributed to Pcf11p (460–469).

3.2. Structural basis for ATP loss by Clp1p in a G135R mutant protein

A number of point mutants of Clp1p have been described in recently published genetic and biochemical studies, all trying to address the role of the Clp1p-bound ATP binding in 3'-end mRNA processing and in CF IA architecture [36–38]. In order to further characterise some of these mutants, we expressed and attempted to purify and to crystallize the Clp1-1p (G135R), the Clp1-6p (G335S), the Clp1-5 (G135R/G335S), Clp1-16p (G135R/L341S) and the Clp1-19p (L341S) in complex with Pcf11p (454–563) (Fig. S2c). Over the various attempts only the Clp1-1p – Pcf11p (454–563) complex led to X-ray diffracting crystals. The Clp1-5p, Clp1-6p and Clp1-16p in complex with Pcf11p (454–563) complex were found unstable and rapidly dissociating during purification. The Clp1-19p–Pcf11p (454–563) was purified to homogeneity but we did

not manage to improve the crystallization conditions and to collect useful X-ray diffraction data.

In the wild type Clp1p, an ATP molecule can be modelled (Fig. 3a and [30]). The interaction network around the ATP has been clearly defined [30]. Due probably a slighted better resolution and/or data quality, a magnesium ion coordinated by the β - γ -phosphate of the ATP and the hydroxyl group of T137 could be modelled. There is probably water molecule around in order to complete the coordination sphere, but none could be added in our refined model (Fig. 3a).

The G135R point mutation leads to a localized rearrangement of the P-loop connecting β -8 to α -2 and involved in ATP phosphate recognition in the wild-type protein (Fig. 3b). In the Clp1-1p, the R135 occupies a position corresponding to K136 position in the wild-type protein (Fig. 3a and b). As such, almost all H-bond interactions undertaken by K136 in the wild type protein are preserved in G135R

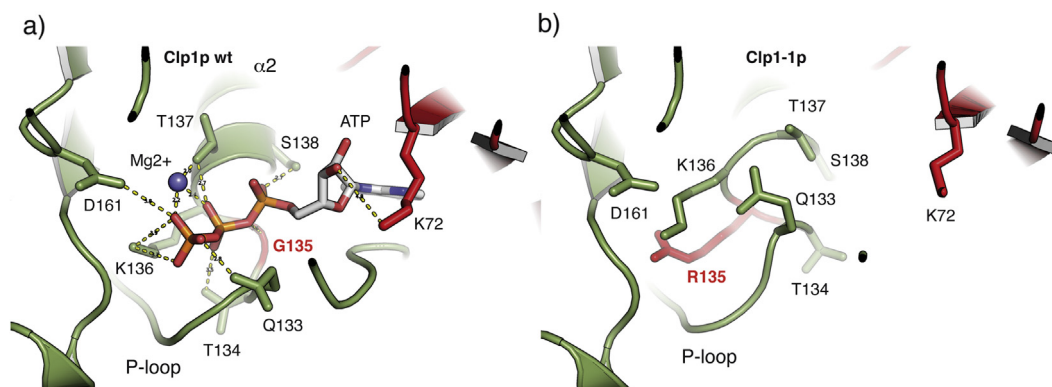


Fig. 3. Structure of the ATP-binding site in the wild-type Clp1p and in the Clp1p G135R (Clp1-1p) point mutant. The two panels are showing Clp1p binding site under the same orientation. Colour coding is similar to Fig. 1 with the exception of the residue 135 shown in red. a) Wild-type Clp1p ligand binding site. Residue side chains involved in ATP-binding are labelled. A Mg²⁺ ion could be modelled in the electron density. The hydrogen-bonding network around the ATP is shown as yellow dotted lines. G135 is coloured in red. b) Clp1-1p (G135R, in red) mutant ligand binding site is remodelled due to the point mutation and no ligand is observed.

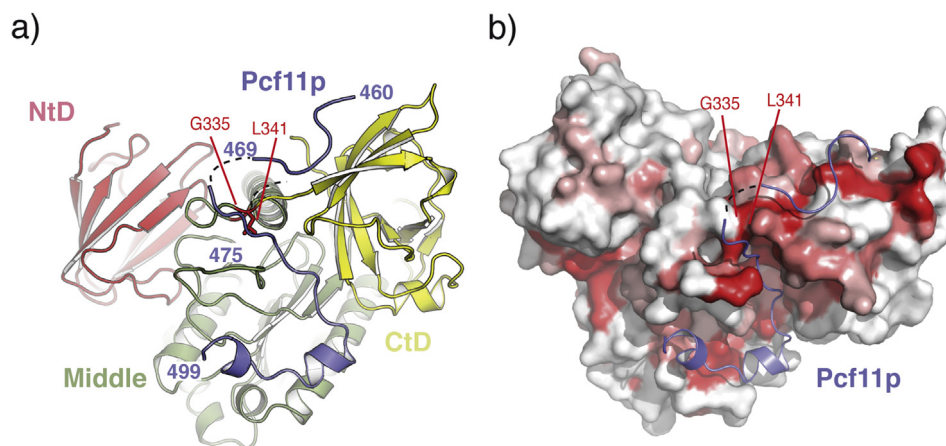


Fig. 4. a) The N-terminal domain (NtD, pink), middle domain (green) and C-terminal domain (CtD, yellow) are shown as described in Noble et al. [30]. Pcf11p is shown in blue. Pcf11p boundaries are indicated. The positions of G335 and L341 mutated in *in vivo* analysis are labelled. b) Sequence conservation at the surface of Clp1p. Yeasts orthologues of Clp1p were aligned and sequence conservation was plotted on the surface of Clp1p as a gradient from white (non-conserved) to red (full conservation). Both structures are shown in the same orientation.

mutant. However, the G135R mutation is accompanied by a flip of the overall main chain between residues 133 to 136 (P-loop) leading to a re-positioning of Q133 and T134. Q133 in the Clp1-1p protein overlaps with the position of the β -phosphate of ATP in the wild-type protein. Altogether, the loop moves approximately 3–4 Å ($C\alpha$ to $C\alpha$ distance) and occludes the ligand-binding pocket. In the absence of nucleotide, the base-binding loop (residues 310–316) is partially disordered. The G135R mutation does not alter the position of D161 and K72; also involved in ATP binding in the wild type Clp1p. No ligand can be seen in the binding pocket. Interestingly, the two Pcf11p binding sites remain unaffected by the G135R point mutation. We conclude that the G135R point mutation does not alter the overall conformation of Clp1p and does not affect Pcf11p binding sites. However, the local rearrangements of the P-loop lead to an inability to bind ATP.

4. Discussion

In *Saccharomyces cerevisiae*, point mutants of Clp1p K136 featuring an alanine [36,38] or asparagine [38] are viable and impair Pcf11p interaction in a pairwise *in vitro* GST pull-down assay [38]. Attempts to reconstitute CF IA with a Clp1p K136A or K136N point mutations failed due to poor stability of the complex [38] suggesting that a mild modification of the structure is sufficient to lead to a drastic change in its biophysical properties. Point mutation of Clp1p K136E leads to a stronger phenotype. This strain is not viable and the recombinant protein harbouring this point mutation does not demonstrate altered ATP binding *in vitro* [36]. The K136E point mutant probably behaves as a knock-out of Clp1p. An endogenous CF IA complex harbouring a Clp1p K136A/T137A double mutation has been purified from yeast [38]. Altogether this demonstrates that ATP binding by Clp1p is not required for viability and that the point mutation alters the overall structure rather than ATP binding. This conclusion is also supported by observations made on the G135R [37] and G131P/Q133D/T134D (recapitulating the human P-loop motif within yeast Clp1p) point mutants [38]. Indeed, both studies lead to the conclusion that Clp1p has lost its ability to bind ATP. Both studies also lead to the conclusion that ATP binding by Clp1p is not needed per se for 3'-end mRNA processing or for transcription termination even though CF IA is required for both processes [37,38]. In human, previous studies have assigned a role for ATP binding by hClp1 [31,33]. The mutation at position K127 (equivalent to K136 in yeast) abolishes the RNA kinase activity

of hClp1, but mice are still viable even though they develop pathologies [33].

The Clp1-1p–Pcf11p (454–563) structure demonstrates that ATP is not needed for *in vitro* interaction between Pcf11p and Clp1p. Even though the ATP binding is largely remodelled, the modification does not alter or propagates towards the Pcf11p binding site. This structure also provides hints for the phenotype observed in strains harbouring G335S or L341S point mutation. Indeed, even though difficult to model, this Pcf11p stretch appears to extend over the area encompassing residue G335 and L341 that were conferring a phenotype to yeast strain when altered. Interestingly, this surface area of Clp1p is conserved through evolution (Fig. 4).

5. Conclusion

In this study, we provide the structural basis for loss of ATP binding by Clp1-1p. We show that in order to accommodate the bulky arginine side chain, the P-loop movement leads to the occlusion of the ATP binding pocket. This local rearrangement however has no consequence for Pcf11p binding interaction. We also report and complete the previous description of the Pcf11p binding interface.

6. Coordinates

Atomic coordinates and structure factors of Clp1p–Pcf11p (454–563) and Clp1-1p–Pcf11p (454–563) complexes have been deposited at the Protein DataBank under the accession numbers 4COB and 4COH respectively.

Acknowledgements

We would like to thank the staff at the ESRF beamlines and the PX1 beamline at SOLEIL for their help during data collection. We thank A. Goldsborough, A. Innis and L. Minvielle-Sébastien for critical reading of the manuscript. We thank L. Minvielle-Sébastien for Clp1p mutants. This work was supported by the ANR-12-BSV8-0016 contract. The Conseil Régional d'Aquitaine and Inserm.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.biochi.2014.01.017>.

References

- [1] S. Millevoi, S. Vagner, Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation, *Nucl. Acids Res.* 38 (2010) 2757–2774.
- [2] N.J. Proudfoot, Ending the message: poly(A) signals then and now, *Genes Dev.* 25 (2011) 1770–1782.
- [3] M. Kim, N.J. Krogan, L. Vasiljeva, O.J. Rando, E. Nedea, J.F. Greenblatt, S. Buratowski, The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II, *Nature* 432 (2004) 517–522.
- [4] S. West, N. Gromak, N.J. Proudfoot, Human 5'→3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites, *Nature* 432 (2004) 522–525.
- [5] C.R. Mandel, Y. Bai, L. Tong, Protein factors in pre-mRNA 3'-end processing, *Cell. Mol. Life Sci.* 65 (2008) 1099–1122.
- [6] Q. Yang, S. Double, Structural biology of poly(A) site definition, Wiley interdisciplinary reviews, *RNA* 2 (2011) 732–747.
- [7] L. Minvielle-Sebastia, P.J. Preker, T. Wiederkehr, Y. Strahm, W. Keller, The major yeast poly(A)-binding protein is associated with cleavage factor IA and functions in premessenger RNA 3'-end formation, *Proc. Natl. Acad. Sci. U. S. A.* 94 (1997) 7897–7902.
- [8] M.M. Kessler, M.F. Henry, E. Shen, J. Zhao, S. Gross, P.A. Silver, C.L. Moore, Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3'-end formation in yeast, *Genes Dev.* 11 (1997) 2545–2556.
- [9] L. Minvielle-Sebastia, K. Beyer, A.M. Krecic, R.E. Hector, M.S. Swanson, W. Keller, Control of cleavage site selection during mRNA 3' end formation by a yeast hnRNP, *Embo J.* 17 (1998) 7454–7468.
- [10] N. Amrani, M. Minet, M. Le Gouar, F. Lacroute, F. Wyers, Yeast Pab1 interacts with Rna15 and participates in the control of the poly(A) tail length in vitro, *Mol. Cell. Biol.* 17 (1997) 3694–3701.
- [11] N. Viphakone, F. Voisinnet-Hakil, L. Minvielle-Sebastia, Molecular dissection of mRNA poly(A) tail length control in yeast, *Nucl. Acids Res.* 36 (2008) 2418–2433.
- [12] E.F. Dunn, C.M. Hammell, C.A. Hodge, C.N. Cole, Yeast poly(A)-binding protein, Pab1, and PAN, a poly(A) nuclease complex recruited by Pab1, connect mRNA biogenesis to export, *Genes Dev.* 19 (2005) 90–103.
- [13] M. Schmid, M.B. Poulsen, P. Olszewski, V. Pelechano, C. Saguez, I. Gupta, L.M. Steinmetz, C. Moore, T.H. Jensen, Rrp6p controls mRNA poly(A) tail length and its decoration with poly(A) binding proteins, *Mol. Cell* 47 (2012) 267–280.
- [14] S. Gross, C. Moore, Five subunits are required for reconstitution of the cleavage and polyadenylation activities of *Saccharomyces cerevisiae* cleavage factor I, *Proc. Natl. Acad. Sci. U. S. A.* 98 (2001) 6080–6085.
- [15] C.M. Hammell, S. Gross, D. Zenklusen, C.V. Heath, F. Stutz, C. Moore, C.N. Cole, Coupling of termination, 3' processing, and mRNA export, *Mol. Cell. Biol.* 22 (2002) 6441–6457.
- [16] D.D. Licatalosi, G. Geiger, M. Minet, S. Schroeder, K. Cilli, J.B. McNeil, D.L. Bentley, Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II, *Mol. Cell* 9 (2002) 1101–1111.
- [17] D. Barilla, B.A. Lee, N.J. Proudfoot, Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in *Saccharomyces cerevisiae*, *Proc. Natl. Acad. Sci. U. S. A.* 98 (2001) 445–450.
- [18] N. Al Husini, P. Kudla, A. Ansari, A role for CF1A 3' end processing complex in promoter-associated transcription, *PLoS Genet.* 9 (2013) e1003722.
- [19] M. Moreno-Morcillo, L. Minvielle-Sebastia, S. Fribourg, C.D. Mackereth, Locked tether formation by cooperative folding of Rna14p monkeytail and Rna15p hinge domains in the yeast CF IA complex, *Structure* 19 (2011) 534–545.
- [20] P. Legrand, N. Pinaud, L. Minvielle-Sebastia, S. Fribourg, The structure of the CstF-77 homodimer provides insights into CstF assembly, *Nucl. Acids Res.* 35 (2007) 4515–4522.
- [21] J.M. Gordon, S. Shikov, J.N. Kuehner, M. Liriano, E. Lee, W. Stafford, M.B. Poulsen, C. Harrison, C. Moore, A. Bohm, Reconstitution of CF IA from overexpressed subunits reveals stoichiometry and provides insights into molecular topology, *Biochemistry* 50 (2011) 10203–10214.
- [22] Y. Bai, T.C. Auferin, C.Y. Chou, G.G. Chang, J.L. Manley, L. Tong, Crystal structure of murine CstF-77: dimeric association and implications for polyadenylation of mRNA precursors, *Mol. Cell* 25 (2007) 863–875.
- [23] T.C. Leeper, X. Qu, C. Lu, C. Moore, G. Varani, Novel protein-protein contacts facilitate mRNA 3'-processing signal recognition by Rna15 and Hrp1, *J. Mol. Biol.* 401 (2010) 334–349.
- [24] Y. Takagaki, J.L. Manley, RNA recognition by the human polyadenylation factor CstF, *Mol. Cell. Biol.* 17 (1997) 3907–3914.
- [25] C. Pancevac, D.C. Goldstone, A. Ramos, I.A. Taylor, Structure of the Rna15 RRM–RNA complex reveals the molecular basis of GU specificity in transcriptional 3'-end processing factors, *Nucl. Acids Res.* 38 (2010) 3119–3132.
- [26] S. Gross, C.L. Moore, Rna15 interaction with the A-rich yeast polyadenylation signal is an essential step in mRNA 3'-end formation, *Mol. Cell. Biol.* 21 (2001) 8045–8055.
- [27] X. Qu, J.M. Perez-Canadillas, S. Agrawal, J. De Baecke, H. Cheng, G. Varani, C. Moore, The C-terminal domains of vertebrate CstF-64 and its yeast orthologue Rna15 form a new structure critical for mRNA 3'-end processing, *J. Biol. Chem.* 282 (2007) 2101–2115.
- [28] M. Sadowski, B. Dichtl, W. Hubner, W. Keller, Independent functions of yeast Pcf11p in pre-mRNA 3' end processing and in transcription termination, *Embo J.* 22 (2003) 2167–2177.
- [29] A. Meinhart, P. Cramer, Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors, *Nature* 430 (2004) 223–226.
- [30] C.G. Noble, B. Beuth, I.A. Taylor, Structure of a nucleotide-bound Clp1–Pcf11 polyadenylation factor, *Nucl. Acids Res.* 35 (2007) 87–99.
- [31] A. Ramirez, S. Shuman, B. Schwer, Human RNA 5'-kinase (hClp1) can function as a tRNA splicing enzyme in vivo, *RNA* 14 (2008) 1737–1745.
- [32] S. Weitzer, J. Martinez, The human RNA kinase hClp1 is active on 3' transfer RNA exons and short interfering RNAs, *Nature* 447 (2007) 222–226.
- [33] T. Hanada, S. Weitzer, B. Mair, C. Bernreuther, B.J. Wainger, J. Ichida, R. Hanada, M. Orthofer, S.J. Cronin, V. Komnenovic, A. Minis, F. Sato, H. Mimata, A. Yoshimura, I. Tamir, J. Rainer, R. Kofler, A. Yaron, K.C. Eggan, C.J. Woolf, M. Glatzel, R. Herbst, J. Martinez, J.M. Penninger, CLP1 links tRNA metabolism to progressive motor-neuron loss, *Nature* 495 (2013) 474–480.
- [34] J. Zhao, M.M. Kessler, C.L. Moore, Cleavage factor II of *Saccharomyces cerevisiae* contains homologues to subunits of the mammalian cleavage/polyadenylation specificity factor and exhibits sequence-specific, ATP-dependent interaction with precursor RNA, *J. Biol. Chem.* 272 (1997) 10831–10838.
- [35] J. Chen, C. Moore, Separation of factors required for cleavage and polyadenylation of yeast pre-mRNA, *Mol. Cell. Biol.* 12 (1992) 3470–3481.
- [36] M.A. Ghazy, J.M. Gordon, S.D. Lee, B.N. Singh, A. Bohm, M. Hampsey, C. Moore, The interaction of Pcf11 and Clp1 is needed for mRNA 3'-end formation and is modulated by amino acids in the ATP-binding site, *Nucl. Acids Res.* 40 (2012) 1214–1225.
- [37] R. Haddad, F. Maurice, N. Viphakone, F. Voisinnet-Hakil, S. Fribourg, L. Minvielle-Sebastia, An essential role for Clp1 in assembly of polyadenylation complex CF IA and Pol II transcription termination, *Nucl. Acids Res.* 40 (2012) 1226–1239.
- [38] S. Holbein, S. Scolia, B. Loll, B.S. Dichtl, W. Hubner, A. Meinhart, B. Dichtl, The P-loop domain of yeast Clp1 mediates interactions between CF IA and CPF factors in pre-mRNA 3' end formation, *PLoS One* 6 (2011) e29139.
- [39] C. Romier, M. Ben Jelloul, S. Albeck, G. Buchwald, D. Busso, P.H. Celie, E. Christodoulou, V. De Marco, S. van Gerwen, P. Knipscheer, J.H. Lebbink, V. Notenboom, A. Poterszman, N. Rochel, S.X. Cohen, T. Unger, J.L. Sussman, D. Moras, T.K. Sixma, A. Perrakis, Co-expression of protein complexes in prokaryotic and eukaryotic hosts: experimental procedures, database tracking and case studies, *Acta Crystallogr. D Biol. Crystallogr.* 62 (2006) 1232–1242.
- [40] W. Kabsch, Integration, scaling, space-group assignment and post-refinement, *Acta Crystallogr. D Biol. Crystallogr.* 66 (2010) 133–144.
- [41] P.D. Adams, P.V. Afonine, G. Bunkoczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger, P.H. Zwart, PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr. D Biol. Crystallogr.* 66 (2010) 213–221.
- [42] P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of Coot, *Acta Crystallogr. D Biol. Crystallogr.* 66 (2010) 486–501.
- [43] O.S. Smart, T.O. Womack, C. Flensburg, P. Keller, W. Paciorek, A. Sharff, C. Vornheim, G. Bricogne, Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER, *Acta Crystallogr. D Biol. Crystallogr.* 68 (2012) 368–380.

Figure S1

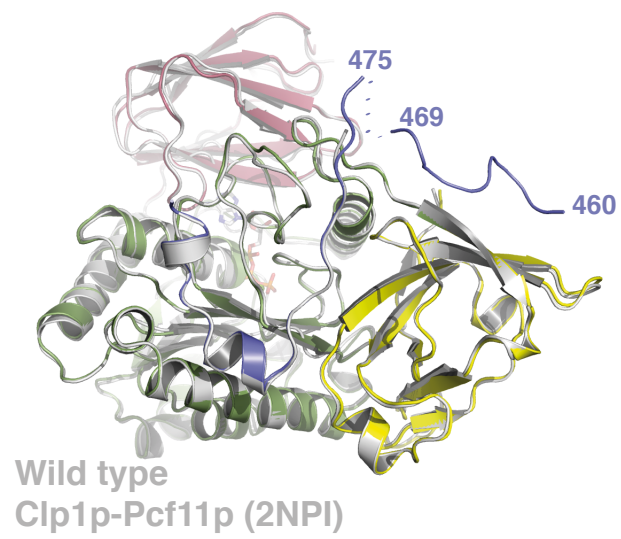
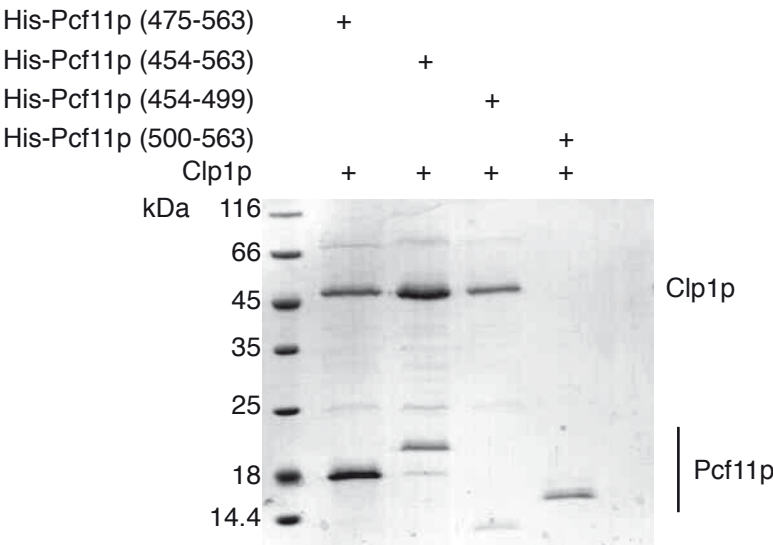


Figure S2

a)



b)



c)

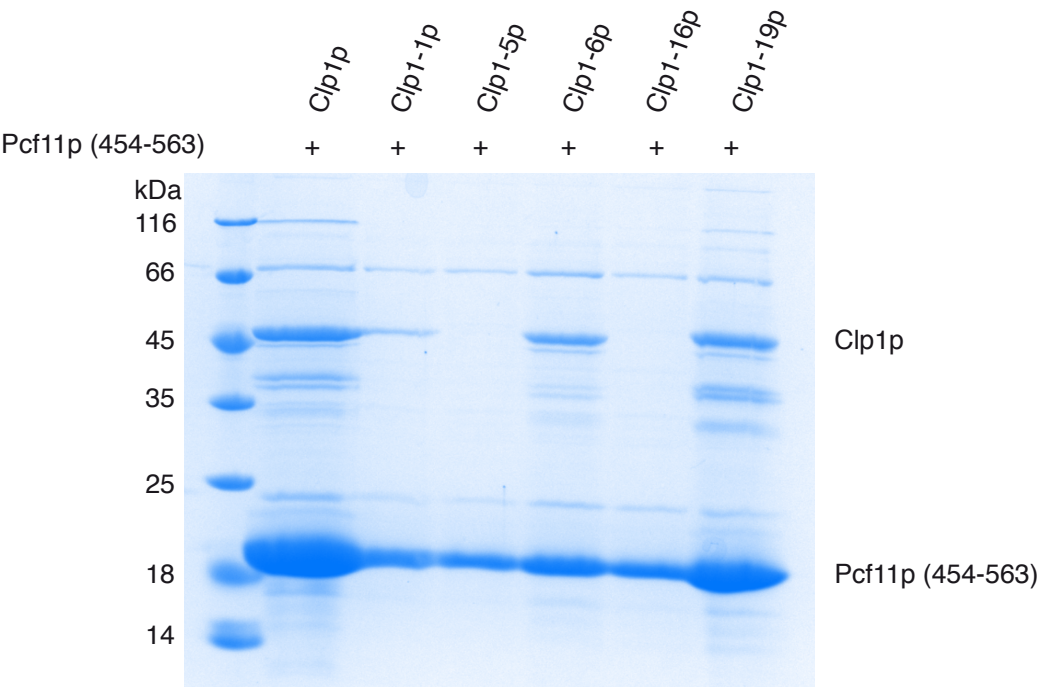


Figure S1:

Clp1-1p-Pcf11p superposition with the 2NPI model [30]. Clp1-1p-Pcf11p (454-563) structure superimposes well on the 2NPI structure (grey) with an overall R.m.s.d. of 0.47 Å on Cα atoms over 428 residues.

Figure S2:

a) Clp1p pull-down experiment using various Pcf11p truncations. On the basis of the structure of Clp1p bound Pcf11p (454-563) [30], various truncations have been designed and tested. The deletion of residues 454-474 does not alter Clp1p binding while the deletion of residues 454-499 abolishes the interaction. Residues 500-563 do not seem to alter Clp1p interaction.

b) Pcf11 sequence conservation around residues 450 to 520 (*S.cerevisiae*). Residues 460-469 (open box) are poorly conserved and found only in *S. cerevisiae*.

c) Pull-down experiments of His-tagged Pcf11p (454-563) with Clp1p variants. Clp1p variants and Pcf11p (454-563) proteins were co-expressed in bacteria and purified over a cobalt-affinity resin. After extensive washes, the proteins were analysed by SDS-PAGE and revealed by Coomassie blue after. The Clp1-5p and Clp1-16p could not be pulled-down by Pcf11p (454-563).

Supplemental Table 1 Crystallographic data collection, processing and refinement statistics.

	Clp1-1p – Pcf11p (454-563)	Clp1p wt –Pcf11p (454-563)
Data collection and processing		
Beamline	Proxima 1	ID23-1
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit cell parameters		
a, b, c (Å)	89.06 95.16 181.19	89.35 95.41 182.14
α, β, γ (°)	90.00 90.00 90.00	90.00 90.00 90.00
Wavelength (Å)	0.9801	0.9763
Resolution limits (Å)	48.0 – 2.70 (2.86-2.70)	40.0 – 2.77 (2.92 – 2.77)
Total reflections	415 071 (65 428)	147 811 (4 615)
Unique reflections	42 917 (6 722)	40 274 (1 376)
Rmerge (%)	9.6 (66.6)	5.2 (39.0)
$I/\sigma I$	21.28 (3.61)	13.9 (1.9)
Completeness (%)	99.7 (98.5)	99.7 (99.77)
Multiplicity	9.46 (9.07)	3.7 (3.4)
Overall B factor (Å ²)	51.98	69.72
CC(1/2)§	0.999 (0.705)	0.998 (0.774)
Refinement statistics		
Resolution range	47.80 – 2.70	40.55 - 2.77
Completeness	99.66	99.59
No. of reflections (working)	43 472	40 190
No. of reflections (test)	2 188	2 784
No. of Non –H atoms	7 311	7 356
Water	25	63
R.m.s. deviations		
Bond lengths (Å)	0.01	0.009
Bond angles (°)	1.24	1.278
Rwork /Rfree (%)	19.06 / 22.30	17.87 / 21.33
Average B factor (Å ²)	59.78	67.87
Protein	56.36	64.08
Water	63.11	71.79
Ramachandran statistics (%)		
Favoured regions	94.04	94.51
Allowed regions	3.82	4.06
Outliers	2.13	1.43
PDB	4C0H	4C0B

Numbers under brackets refer to the highest resolution shell.

§ According to Karplus et al. [44], the highest resolutions are 2.77 Å and 2.70 Å resolution using the criteria $CC_{1/2} > 0.7$ and $I/\sigma I > 2.0$.

4. Conclusions et discussion

Si la maturation de l'extrémité 3' de l'ARN messager est considérée comme un mécanisme essentiel de la biologie cellulaire, affectant de nombreux aspects du devenir de l'ARN, le nombre considérable de facteurs nécessaires reste impressionnant au vu de la relative simplicité des réactions catalysées. Les mécanismes biochimiques du clivage et de la polyadénylation en tant qu'activités enzymatiques sont de mieux en mieux caractérisés, notamment grâce à l'identification des protéines portant ces activités, mais l'étape cruciale qu'est la sélection du site de polyadénylation qui va être maturé est encore mal connue. Les caractéristiques canoniques du site de polyadénylation sont bien décrites, aussi bien chez les mammifères que chez les levures ou les plantes. Cependant, en cas de compétition entre plusieurs sites, *i. e.* la polyadénylation alternative, ces caractéristiques communes et conservées ne suffisent pas forcément à expliquer le choix du site.

Le Cleavage Factor IA (CF IA), bien que ne portant aucune activité enzymatique, joue un rôle central dans la maturation. Précisément dans la sélection du site de maturation, en interagissant avec le Positioning Element sur le pré-ARNm. Parvenir à caractériser structuralement, ou tout au moins topologiquement, cet intermédiaire clé du positionnement des enzymes sur leur substrat est donc une condition *sine qua none* de la compréhension des mécanismes de la maturation et pourrait d'ores et déjà expliquer une partie de la polyadénylation alternative.

Durant ce travail, ma stratégie a consisté à mettre au point un outil adapté, le CF IA recombinant, puis à exploiter au maximum les possibilités offertes par ce matériel.

Dans un premier temps, je me suis attaché à établir le caractère fonctionnel de mon modèle. En effet, aucun travail sur un CF IA produit par co-purification n'a jamais été proposé dans la littérature, et il m'a semblé indispensable de m'assurer que nous conservions les caractéristiques fonctionnelles pour accorder un plein crédit aux informations obtenues par ce modèle. Nous avons donc testé, avec succès, la capacité du complexe recombinant à compléter un essai *in vitro* dépourvu de CF IA natif. Au-delà de la confirmation de notre modèle et de nos procédés de purification, le fait qu'un facteur produit chez la bactérie soit fonctionnel *in vitro* semble indiquer que le complexe ne nécessite pas de modifications post-

traductionnelles spécifiques pour jouer son rôle dans la maturation, bien que cela n'exclue pas des régulations *in vivo*, par phosphorylation par exemple.

Dans un second temps, j'ai exploité l'avantage de la co-expression et de la co-purification du complexe pour cartographier plus finement la zone d'interaction entre Pcf11p et Rna14p-Rna15p sur Pcf11p. Cette approche a permis de réduire le fragment nécessaire et suffisant pour l'interaction à la région comprise entre les résidus 331 et 400, là où une précédente étude de double hybride avait déterminé les résidus 288 et 400 comme bornes du domaine. Cette information, et la connaissance de la région d'interaction avec Clp1p, m'a alors permise de définir ce que je nomme le Pcf11p minimal, c'est-à-dire nécessaire et suffisant pour l'association du CF IA. L'idée derrière la conception de cette construction est d'exploiter au maximum l'indépendance fonctionnelle des domaines de Pcf11p (Sadowski, Dichtl et al. 2003) et de réduire l'importance des régions avec une tendance au désordre.

Malgré ce travail et cette optimisation de mon modèle, les essais de cristallogénèse sur le complexe ne se sont pas révélés concluants et n'ont pas permis l'identification de conditions favorables à une stabilisation et une organisation du facteur dans une maille cristalline. Seul un fragment de dégradation de Rna14kl, déjà obtenu par des essais de protéolyse (Bai, Auperin et al. 2007) et également retrouvé lors d'essais avec Rna14kl-Rna15kl (Paulson and Tong 2012), a pu être obtenu. Cet échec, symptôme d'un complexe portant encore trop de régions flexibles, est finalement en accord avec la fonction du CF IA. La principale mobilité subsistant dans notre facteur portant le Pcf11 minimal est en effet apportée par Rna15, une protéine portant trois domaines caractérisés structuralement reliés par des séquences très probablement flexibles. Cette flexibilité permettrait en effet aux RRM de disposer de la liberté nécessaire pour reconnaître spécifiquement le site pA, tout en gardant une marge d'adaptation à la multitude de variations existant au sein du génome. Travailler avec un CF IA fonctionnel nous condamne donc à nous accommoder de cette contrainte.

Pour ce faire, nous avons donc développé deux autres approches, la microscopie électronique et la diffusion des rayons X aux petits angles. Mais la trop grande liberté apportée par la mobilité dans le cas des modélisations SAXS et l'absence de clichés exploitables pour des calculs d'enveloppe en microscopie électronique nous ont à nouveau conduits dans une impasse, malgré notre opiniâtreté. Ces approches, et notamment le SAXS, confirmant

cependant la qualité de nos échantillons mais aussi la mobilité liée à Rna15, nous ont encouragées à exploiter notre modèle d'une autre manière.

Ceci s'est tout d'abord concrétisé par l'analyse du facteur en spectrométrie de masse en condition natives, un champ vierge de l'étude du CF IA. Notre objectif est alors d'obtenir des informations topologiques, et notamment sur la stœchiométrie du complexe. De nombreuses analyses préliminaires nous ont finalement autorisées à interpréter des données acquises sur le complexe. Une ambiguïté est cependant toujours présente, deux stœchiométries pouvant expliquer les résultats obtenus à cause d'une dégradation importante de Pcf11p. Cette stœchiométrie soulève des questions extrêmement intéressantes, mais la seule chose à faire pour le moment est de renouveler les analyses. Il m'apparaît très tentant à ce stade d'essayer d'employer le Pcf11p minimal afin de limiter –voire supprimer– la dégradation.

Finalement, afin de réguler la mobilité intrinsèque du complexe, nous avons également opté pour une « complexification » du modèle, en y ajoutant les principaux co-facteurs que sont l'ARN et le CF IB (Hrp1), et ainsi reformer le CF I. J'ai tout d'abord voulu caractériser le complexe CF I lié à son partenaire ARN, afin d'aborder la cristallogénèse avec la stratégie la plus adaptée. Les approches de SPR et de retard sur gel se passent bien avec Hrp1p et l'ARN, me permettant d'obtenir une constante d'association similaire à celle décrite dans la littérature. Cependant, il m'a malheureusement été impossible d'obtenir la moindre donnée sur le CF IA couplé à l'ARN, et *a fortiori* pour le CF I couplé à l'ARN, avec ces approches. Comme je sais, grâce aux essais de clivage et de polyadénylation, que mon complexe recombinant est fonctionnel et donc capable de s'associer avec ses partenaires, j'ai tout de même lancé des essais de cristallogénèse. Aucune des conditions explorée ne s'est révélée propice au développement de cristaux, mais, bien que je n'aie pas eu le temps de le faire moi-même, je pense qu'il pourrait être judicieux de lancer de nouveaux essais avec plus de matériel ou peut être un large excès d'un des partenaires.

Finalement, en parallèle de ces approches globales du complexe CF IA, je me suis également intéressé à la présence de doigts de zinc très conservés mais hypothétiques sur Pcf11p, dont nous démontrons ici la présence. Nous n'avons pour le moment pas d'informations plus précises sur leurs fonctions. Des tests fonctionnels réalisés avec des mutants apporteront bientôt un début de réponse, et des essais pour résoudre leur structure

sont également en cours. Les doigts de zinc sont des motifs très courants, et plusieurs facteurs en portent au sein de la machinerie de maturation 3'. Il est improbable qu'ils soient impliqués dans une interaction avec des acides désoxyribonucléiques, aucune des protéines du CF I n'est connue pour lier l'ADN et seules Hrp1p et Rna15p interagissent avec l'ARN (Gross and Moore 2001). Il est donc probable que leur conservation soit liée à une interaction avec un des multiples partenaires protéique de Pcf11p. Il est probable que cette interaction implique un facteur hors de CF IA puisque la mutation des résidus permettant la chélation du n'affecte l'architecture de CF IA.

Je me suis également intéressé à l'ATP lié de manière conservée par Clp1p, sans qu'aucune activité d'hydrolyse ne soit jamais mise en évidence ailleurs que chez les eucaryotes supérieurs et les archées. Mais pour autant, des mutants défectueux pour la liaison à l'ATP ne sont pas viables pour la maturation 3'. Notre publication met en évidence, grâce à la résolution de la structure du mutant G135R, que l'absence d'ATP entraîne une perturbation de la structure de la P-Loop et un remodelage du site d'interaction. Nos différents essais nous ont également permis d'obtenir des cristaux avec la protéine sauvage, dont l'analyse a révélé une extension de l'interaction entre Pcf11p et Clp1p. La région d'interaction sur Clp1p est extrêmement conservée entre les levures, alors que la séquence identifiée sur Pcf11p n'est conservée que chez *Saccharomyces cerevisiae*. Considérant le fait que cette interaction n'ait pas été identifiée précédemment, cela laisse penser que nous sommes en présence d'une interaction conservée, mais non essentielle.

A mon avis, il est maintenant indispensable de caractériser la zone d'interaction de Rna14p avec Pcf11p. Il s'agit en effet de la dernière inconnue dans l'assemblage du CF IA, et je pense que c'est également la cause de l'asymétrie au sein du complexe. Caractériser cette interaction devrait donc nous permettre d'avancer, voire de répondre, à la question de la stœchiométrie et par extension de la reconnaissance et de la spécificité à l'ARN. La réponse à cette question viendra aussi de la spectrométrie de masse, notamment en faisant des analyses avec un CF IA portant la construction de Pcf11p minimale. La caractérisation de la zone d'interaction pourrait également nous autoriser à essayer de nouvelles modélisations du CF IA à partir des données SAXS, en ajoutant un point d'ancrage supplémentaire au grand flou qu'est pour le moment le domaine central de Pcf11p.

5. Matériel et méthodes

5.1. Biologie moléculaire

5.1.1. Bactéries utilisées

Durant ma thèse, j'ai exclusivement utilisé des souches optimisées d'*Escherichia coli*. Les principales souches employées pour le clonage sont les DH5α et les XL1 blue dont les caractéristiques majeures sont l'inactivation de l'endonucléase A1 et du système hsd digérant l'ADN non méthylé. Pour l'expression, les BL21 (DE3) seront quasiment exclusivement utilisées. Leurs caractéristiques majeures permettent l'expression avec un système T7 ainsi que d'être déficientes en protéases Lon et OmpT.

Principales souches utilisées durant ma thèse parmi celles disponibles au laboratoire :

Souche	Génotype
XL1 blue	<i>recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac</i> [F' proAB lacI ^q ΔM15 Tn10 (Tet ^r)]
DH5α	F ⁻ Φ80 <i>lacZ</i> ΔM15 Δ(<i>lacZYA-argF</i>) U169 <i>recA1 endA1 hsdR17</i> (rK ⁻ , mK ⁺) <i>phoA supE44 λ- thi-1 gyrA96 relA1</i>
BL21 (DE3)	<i>E. coli B F⁻ ompT hsdS_B(r_B⁻ m_B⁻) gal dcm</i>
Rosetta	<i>F⁻ ompT hsdS_B(r_B⁻ m_B⁻) gal dcm pRARE² (Cam^R)</i>
BL21 Gold pLysS	<i>E. coli B F⁻ ompT hsdS_B(r_B⁻ m_B⁻) gal dcm + Tetr gal endA Hte</i>

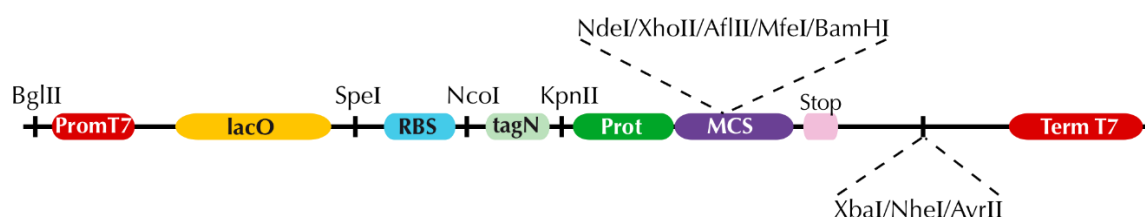
5.1.2. Plasmides utilisés

Au laboratoire, nous utilisons des plasmides avec le squelette pET permettant l'expression inducible en système bactérien. Ces plasmides portent un gène de résistance sous la dépendance d'un promoteur ubiquitaire et une cassette d'expression constituée d'un site de liaison du ribosome (RBS) et du multi site de clonage MCN sous la dépendance du promoteur de l'opéron lactose, et donc inducible à l'IPTG. La région MCN permet l'introduction d'une étiquette et de la séquence cible de la protéase TEV en N ou en C terminal

du cadre de lecture. Cette construction nous permet alors d'insérer notre gène d'intérêt, où il sera exprimé en fusion avec le site TEV et l'étiquette.

Etiquettes utilisées au laboratoire :

Etiquette	Séquence	Masse (kDa)
His-TEV	MGSSHHHHHHSSGTGSGENLYFQG	2,6
His-GST-TEV	MSGSHHHHHHSGSGSMAPILGYWKIKGLVQPTRLLEYLEEKYEEH LYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIRYIADK HNMLGGCPKERAIEISMLEGAVLDIRYGVSRAYSQDFETLKVDFLSKL PEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCL DAFPKLVCFKKRERIAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPP KSDGTGSGENLYFQG	28,4
His-GB1-TEV	MSGSHHHHHHSGSGSMGSSMKQYKLILNGKTLKGETTTEAVDAAT AEKVFKQYANDNGVDGEWYDDATKTFTVTESSGTGSGENLYFQG	6,2
His-SUMO-TEV	MSGSHHHHHHSGSGSMSDSEVNQEAKPEVKPEVKPETHINLKVSD GSSEIFFKIKKTTPLRRLMEAFKRQKGEMDSLRLFLYDGIRIQADQTP EDLDMEDNDIIEAHREQIGGATYGTGSGENLYFQG	7,3



Les séquences complètes des vecteurs sont disponibles par le lien ci-dessous :

http://archive.igbmc.fr/recherche/Sup_papers/Cavarelli/pET-MCN/pET-MCN.html

5.1.3. Extraction d'ADN génomique de *Kluyveromyces lactis*

Des levures *Kluyveromyces lactis*, souche PM6-7A, généreusement cédées par Marc Lemaire, sont étalées sur une boîte YPDA et mises à pousser à 30°C. Les colonies sont alors remises en suspension dans 1 M Sorbitol/0.1 M EDTA et centrifugées. Le culot est repris dans

300 µl de tampon 1 M Sorbitol/0,1 M EDTA + 14 mM β-MercaptoEthanol + 0,4 mg/ml zymolyase et laissé à incuber à 37°C pour environ 30 min. L'action de la zymolyase peut être contrôlée par une observation au microscope. Le lysat est alors centrifugé 30 min à vitesse maximale puis le culot est repris dans 400 µl de tampon Tris/EDTA. On y ajoute alors 100µl de tampon de lyse (0,5 M EDTA, pH 8.0 ; 2 M Tris, pH 8.0 ; 10% SDS) et on incube à 65°C pour 30 min. 100 µl de 5 M KoAC sont ensuite ajoutés à la réaction qui est homogénéisée et laissée à 4°C une heure ou sur la nuit. 50 µl de chloroforme sont alors ajoutés, et la solution est centrifugée à vitesse maximale à 4°C pour 15 min puis le surnageant est transféré dans un tube neuf. Un volume équivalent d'éthanol est mixé avec le surnageant et le tout est centrifugé à nouveau 10 min à 4°C à vitesse maximale. Le culot est lavé avec de l'éthanol à 50%, laissé à sécher 5 min à température ambiante puis repris dans 400 µl de Tris/EDTA. 0,025 µg de RNase A sont ajoutés à la solution qui est mise à incuber 30 min à 37°C. La phase aqueuse est extraite avec un volume équivalent de phénol : chloroforme dans un premier temps, puis avec un volume équivalent de chloroforme. L'ADN est alors précipité à -20°C pendant 20 min après l'ajout de 120 µl de 7,5 M NH₄OAc et de 1 ml de 100% EtOH. L'échantillon est alors centrifugé à vitesse maximale pour 15 min, à 4°C, et le culot est lavé à l'éthanol 70%, séché à l'air libre et repris dans 50 µl de tampon Tris/EDTA.

Un aliquot de *Kluyveromyces lactis* est repris dans 50% YPDA et 15% glycérol pour être congelé dans de l'azote liquide et conservé à -80°C.

5.1.4. Clonage

Les clonages sont réalisés par LIC (Ligase Independent Cloning), un protocole complet et détaillé est disponible au téléchargement :

<http://www.iecb.u->

PCR product generation (insert)

His-LIC-ORF (1+)

NdeI

ctttacttccagggccatag- ORF

LIC-ORF (1+)

NdeI

ctttaagaaggagatatacatag- ORF

LIC-ORF (-)

BamHI

tctagactattaggatcc -ORF

Figure 59 : séquences des amorces utilisées en LIC.

bordeaux.fr/teams/FRIBOURG/Fribourg/Protocol_files/SLIC-protocol.pdf

Les grandes lignes sont présentées ici. Les séquences codantes correspondantes aux protéines à exprimer sont amplifiées par PCR à partir de l'ADN génomique grâce à des amorces s'hybridant, pour 18 nucléotides sur la séquence codante, et pour 18 nucléotides sur la séquence du plasmide receveur (Figure 59). Les produits de PCR sont alors purifiés par migration sur des gels d'agarose colorés au SYBR green.

Le plasmide receveur est ouvert de manière à ce que l'insert puisse s'hybrider de part et d'autre du site de restriction et déphosphorylé à l'aide de la rSAP (recombinant Shrimp Alkaline Phosphatase). Le vecteur est alors également purifié sur un gel d'agarose au SYBR green et purifié avec le kit Qiagen QiaexII.

La réaction de LIC est alors lancée, 10-15 minute à 22°C comme suit :

NEB 2 10X buffer	1 µl
BSA 10 mg/ml	1 µl
Vecteur purifié	100 ng
Insert purifié	200-300 ng
T4 DNA Pol	0,2 µl
Eau qsp	10 µl

La réaction est ensuite laissée dans la glace et transformé dans des DH5α chimio-compétentes (Inoue Protocol). Après sélection sur boîte, les ADN plasmidiques sont extraits des clones résistants et confirmés par digestion avec des enzymes de restriction pertinentes et par séquençage.

5.1.5. Expression des protéines recombinantes

Des bactéries *E.coli* BL21 DE3 électro-compétentes sont transformées avec jusqu'à quatre plasmides codants pour les protéines d'intérêt puis étalées sur milieu sélectif.

Pour une expression à petite échelle, afin de tester l'expression ou l'interaction, un tube de 50ml contenant 10 ml de LB sélectif seraensemencé avec un clone résistant et laissé à 37°C, 220 rpm jusqu'à turbidité (environ 5h-7h). La mini culture est alors passée à 15°C, 220 rpm pour une heure afin de limiter la multiplication bactérienne et l'expression est induite sur la nuit avec 1 mM d'IPTG. Après culottage et élimination du milieu, les bactéries sont lysées au sonicateur (5s, 6W output, 4°C) dans du tampon de lyse (PBS 1,5X, Acétate Mg²⁺ 1mM, NP-

40 0,1%, Imidazole 20mM, Glycérol 10%). Un échantillon de lysat peut être conservé pour vérifier que les protéines ne sont pas dans la fraction insoluble. Les débris cellulaires sont éliminés en centrifugeant (10 min, 13000 rpm, 4°C) et le surnageant contenant les protéines solubles est mis à incuber avec 30µl de résine Nickel (Sigma) (1h, 4°C, agitation). Un échantillon de surnageant peut également être conservé pour vérifier que les protéines sont bien dans la fraction soluble. Les billes sont alors lavées par trois cycles de culottage, élimination du surnageant et resuspension à basse vitesse (2000 rpm) dans du tampon de lyse successifs. Après le dernier lavage, l'échantillon est dénaturé avec du tampon de charge SDS () pour être analysé sur gel d'acrylamide.

Pour une expression à grande échelle, une préculture sélective de bactéries transformées est faite sur la nuit pour ensemer un litre de milieu TB dans des erlenmeyers avec déflecteurs. Les cultures sont mises à 37°C, 220 rpm jusqu'à une densité optique de 1.2, et sont alors passées à 15°C, 220 rpm. Une fois la température redescendue, la culture est induite sur la nuit avec 0,25 mM IPTG. Le lendemain, les bactéries sont culottées et le culot est conservé à -20°C.

5.1.6. Gel retard

Le protocole de retard sur gel est inspiré de celui décrit par Noble et al. en 2004 (Noble, Walker et al. 2004). Tout d'abord, 3 µmol de sonde sont marqués à l'ATP 8P³² par le T4 kinase et purifiés sur colonne Microspin G25 selon les instructions du fabricant. La réaction d'interaction est ensuite effectuée en incubant 20 min à 4°C des quantités croissantes de protéine (0,02 à 0,4 µmol) avec une quantité fixe (0.33 µmol) d'ARN radiomarké dans du tampon d'interaction 1X (10% glycerol, 5 mM MgCl₂, 4 mM DTT, 40 mM Tris-HCl, pH 8.0). Les réactions sont alors résolues sur un gel de polyacrylamide natif 4,5% (TBE, 35% acrylamide/Bis-acrylamide), 100 V, 3h. Après fixation, les éléments radioactifs sont révélés par autoradiographie.

5.1.7. Purification de l'ARN polymérase du phage T7 P266L

Des bactéries BL21 (DE3) sont transformées avec le plasmide d'expression pBH161 contenant le gène codant pour l'ARN polymérase T7 P266L. Les clones résistants sont mis à pousser en milieu LB sélectif à 37°C jusqu'à ce que OD₆₀₀=0.5. L'expression est alors induite

par 0.25 mM d'IPTG pendant 4h à 37°C. Les bactéries sont récupérées par centrifugation et les culots peuvent être congelés et stockés. Pour la purification, le culot est repris dans 20 mM Tris-HCl pH 8.0, 500 mM NaCl pour un volume final de 100 ml et lysé par deux passages dans l'émulsifiant à 1000 psi. Le matériel insoluble est culoté par une centrifugation à 50 000g, 1h, 4°C dans un rotor Ti45. Le surnageant filtré (0,22 µm) est chargé sur des colonnes HisTrap équilibrées avec 97,5% de tampon A (20 mM Tris-HCl pH 8,0 ; 500 mM NaCl) / 2,5% tampon B (20 mM Tris-HCl pH 8,0 ; 500 mM NaCl ; 200 mM Imidazole pH 8,0). La résine est alors lavée avec 90% de tampon A / 10% tampon B puis la polymérase est éluée avec 50% tampon A / 50% tampon B. Les fractions du pic sont alors regroupées et diluées d'un facteur 3 dans le tampon de dilution (50 mM Hepes-KOH pH 7,0 ; 2 mM EDTA ; 5% Glycerol ; 2 mM DTT) et chargées sur une colonne mono-S équilibrée en Tampon bas sel (50 mM Hepes-KOH pH 7,0 ; 75 mM NaCl ; 2 mM EDTA ; 5% Glycerol ; 2 mM DTT). L'éluion est effectuée avec un gradient 0-100% de tampon haut sels (50 mM Hepes-KOH pH 7,0 ; 1 M NaCl ; 2 mM EDTA ; 5% Glycerol ; 2 mM DTT) sur 200 ml, la polymérase se trouvant dans le pic du milieu. Après analyse sur gel d'acrylamide, les fractions d'intérêt sont concentrées jusqu'à obtenir 2 ml qui seront chargées sur une colonne Superdex 200 HiLoad 16/60 équilibrée dans du tampon de stockage (10 mM Tris-HCl pH 7,8 ; 200 mM NaCl ; 2 mM DTT). Après éluion, les fractions d'intérêt sont regroupées et concentrées jusqu'à 40 mg/ml, diluées dans 50% glycérol, aliquotées et conservées à -80°C.

5.1.8. Transcription in vitro

Les matrices de transcription sont générées par PCR et purifiées sur gel pour être reprises dans de l'eau traitée au DEPC. La réaction de transcription est préparée selon le tableau suivant. Le tampon Robomax est constitué de HEPES-KOH pH 7,5 400 mM ; MgCl₂ 120 mM ; DTT 200 mM ; Spermidine 10 mM dilués dans l'eau DEPC.

	Stock	Final
Matrice		5ng.µl
Tampon Ribomax	5X	1X
NTP pH7,5	25mM	1mM
T7 RNA Pol		50-100 µg/µl

DEPC MilliQ H ₂ O		qsp
------------------------------	--	-----

La réaction est incubée 4h dans un incubateur à 37°C puis centrifugée 10 min à 13000 rpm, 4°C. Le surnageant est dosé, analysé sur gel TBE/Urée et conservé à -80°C. L'ARN peut également être précipité avec 0,3 M de NaAc et deux volumes d'éthanol afin d'être repris dans de l'eau DEPC.

5.1.9. Western Blot

Les échantillons sont migrés sur un gel d'acrylamide dénaturant puis transférés sur membrane de nitrocellulose avec un système semi-sec (200 mA, 45 min) ou un système Trans Blot Turbo (Biorad) (un Minigel = 1,3A, 25V, 7min) dans du tampon de transfert (Tris-HCl 25 mM, glycine 192 mM). La membrane est alors saturée dans du PBS ; 0,1% Tween ; 5% lait pendant une heure puis incubée avec du PBS ; 0,05% Tween ; 5% lait avec l'anticorps primaire selon la dilution indiquée dans le tableau ci-dessous pendant une heure à température ambiante. Après trois lavages PBS ; 0,1% Tween, la membrane est incubée avec l'anticorps secondaire dilué dans du PBS ; 0,05% Tween ; 5% lait pendant une heure. La membrane est à nouveau lavée pour être révélée avec le kit SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific).

L'anticorps secondaire d'âne anti lapin couplé à la peroxydase est disponible chez GE Healthcare

Protéine révélée	Anticorps primaire	Anticorps secondaire
Rna15p	Polyclonal Lapin (1/10000)	1/4000
Rna14p	Polyclonal Lapin (1/3000)	1/4000
Clp1p	Polyclonal Lapin (1/1000)	1/4000
Pcf11p	Polyclonal Lapin (1/3000)	1/4000

5.1.10. Protéolyse ménagée

Des quantités décroissantes de CF IA purifié sont incubées avec une quantité fixe de protéase par dilution en série pour une heure à 4°C. La réaction est ensuite résolue par migration sur un gel d'acrylamide dénaturant.

Liste des protéases utilisées :

Protéase	Fournisseur
Chymotrypsine	Fluka
Trypsine	Euromedex
Thermolysine	Fluka
Thrombine	mpBio
Pepsine	Euromedex

5.2. Biologie

5.2.1. Test en goutte et thermosensibilité

La croissance et la thermosensibilité des mutants obtenus est étudiée sur milieu Yeast extract Peptone Dextrose supplémenté Adénine - YPD(A). Une anse de cellules est prélevée de cultures sur YPD(A) et reprise dans 50µL d'H₂O milliQ stérile. Après dilutions sérielles, 5µL de chaque dilutions sont déposée sur milieu YPD(A) et incubés aux différentes températures pendant 24 à 48h.

5.2.1. Test de croissance

Les retards de croissance observés lors des tests de thermosensibilité sont étudiés en milieu liquide, à 30°C ; afin de pouvoir calculer le temps de génération. 200mL de milieu YPD(A) sont inoculés à partir d'une pré-culture sur la nuit pour obtenir une DO_{600nm} de 0,05. Les cultures sont incubées à 30°C, 160rpm durant 24h à 48h et la DO_{600nm} est évalué toutes les heures à partir de 1mL de culture.

5.2.2. Tests d'activité

- Production d'ARN radioactif *in vitro*

La maturation en 3' des ARN messagers est étudiée par reconstitution de cette activité *in vitro*. Un ARN synthétique est fourni comme substrat lors de la réaction. Deux types d'ARN peuvent être utilisés : CYC1 ou CYC1pre ; correspondant à la région 3' non traduite de l'ARN du gène CYC1, codant pour l'iso-1-cytochrome c. Les matrices d'ADN employées sont les plasmides pG4-CYC1 (promoteur T7 – Fragment +351/+588 – Minvielle Sebastia et al, 1994) et pG4-CYC1pre (promoteur T7 – Fragment +351/+505 – Preker et al, 1995) linéarisés respectivement par EcoRI et NdeI. La synthèse est catalysée par l'ARN polymérase du bactériophage T7 (25 unités ; Proméga) dans son tampon ; auquel est ajouté : 50ng/μL ADN matrice, 0,01% TX-100, 0,5mM ATP, 0,5mM CTP, 0,1mM UTP, 0,1mM GTP, 1mM m7GpppG, 2 unités/μL d'inhibiteur de RNase. Le mélange réactionnel contient 1μL d'[α32P]-UTP. Après incubation 2h à 37°C, l'ARN est précipité par l'ajout de 500μL d'éthanol 100%, 50μL d'acétate d'ammonium et 0,5μL de glycogène (Euromedex, 20mg/mL). L'ARN est récupéré par centrifugation, puis lavé dans de l'éthanol 70% et séché pour être repris dans 20μL de tampon de charge (98% formamide, 10mM EDTA, 0,125% de Xylène cyanole FF et 0,125% de bleu de bromophénol) et dénaturé 2 minutes à 95°C.

- Préparation et chargement d'un gel 6% polyacrylamide-urée

Les ARN synthétiques sont séparés par électrophorèse sur gel 6% polyacrylamide (acrylamide/bisacrylamide 19 :1) / TBE 1X / 8,3M urée. Les plaques de 20 cm x 20 cm sont séparées par des espaceurs de 0,75 mm. Le tampon de migration est composé de 45mM Tris-borate, 1mM EDTA (TBE 1X – Biosolve Chemicals). 20mL de solution d'acrylamide (+ 120μL APS 10% et 20μL TEMED) sont nécessaires pour former un gel. Ce dernier est ensuite soumis à une pré-migration à 900V pendant 20 minutes. Les ARN synthétisés sont chargés dans des puits larges (1,5cm) et la migration est effectuée à 900V pendant 45 minutes environ.

- Purification des ARN synthétisés

L'ARN est repéré sur le gel par auto-radiographie : un film Kodak BioMax MS est exposé 1 minute. La bande est découpée et l'ARN est élué dans 400μL de tampon d'éluion (0,75M acétate ammonium, 10mM MgAc, 0,1mM EDTA et 0,1% SDS) contenant 4μL d'un mélange phénol/chloroforme/alcool isoamylique (25/24/1) et 0,5μL de glycogène (Boehringer, 20mg/mL). Après incubation sur la nuit à 37°C, le surnageant est mélangé à 400μL d'un

mélange phénol/chloroforme/alcool isoamylique (25/24/1). Après centrifugation, la phase aqueuse est mélangée à 1mL d'éthanol 100% froid et les ARN synthétisés sont stockés à -20°C.

- Calculs

Ces mesures sont réalisées deux fois.

Activité totale (A) :

1μL de mélange réactionnel initial est prélevé et dilué au centième. On mesure l'activité X (en coups par minutes, cpm), présent dans 5μL de cette dilution.

$$A = (X/5) \cdot 100 \cdot 20 = X \cdot 400$$

Activité spécifique (B) :

Le mélange réactionnel contient $2 \cdot 10^{-9}$ mol UTP. L'activité spécifique B du mélange réactionnel est exprimée en cpm / pmol UTP

$$B = (A/2 \cdot 10^{-9}) \cdot 10^{-12} = A \cdot 5 \cdot 10^{-4}$$

Activité incorporée (C) :

L'activité de 10μL (des 1400μL final) de suspension d'ARN est mesurée.

$$C = (1400/10) \cdot Y = 140 \cdot Y$$

Quantité d'UTP incorporé (D) :

Elle est exprimée en pmol d'UTP.

$$D = C / B$$

Concentration de la suspension d'ARN (E) :

L'ARN CYC1 contient 97 résidus Uracile et CYC1-pré en contient 61. La quantité (notée E) d'ARN dans la suspension finale, exprimée en pmol d'ARN, est alors égale à :

- pour CYC1 : $E = (1/97) \cdot D$

- pour CYC1-pré $E = (1/61) \cdot D$

La concentration (F) de la suspension d'ARN est exprimée en fmol/ μ L

$$F = E / 1,4$$

- Réaction de reconstitution des activités de maturation en 3' des ARN messagers

Le mélange réactionnel (cf Tableau) utilisé et l'ARN substrat choisi permettent d'étudier les activités de clivage seul, de polyadénylation seule ou de clivage et de polyadénylation couplées. La présence de 2mM d'EDTA permet d'analyser le clivage seul d'un ARN substrat CYC1. La présence de 2mM de MgAc permet d'analyser le clivage et la polyadénylation d'un ARN substrat CYC1, ou la polyadénylation de l'ARN substrat pré-clivé - CYC1-pré. Le mélange réactionnel est incubé 1 heure à 30°C. 75 μ L d'un tampon contenant 1 μ L de protéinase K (20mg/mL) sont ajoutés et le mélange est incubé 30 minutes à 42°C. Cette étape permet la dégradation des protéines et la libération des ARN, qui peuvent être récupérés par précipitation par ajout de 2,5 volumes d'éthanol 100% froid. Après centrifugation, les culots d'ARN sont lavés par 1mL d'éthanol 70%, séchés puis repris dans 6 μ L de tampon de charge et dénaturés 2 minutes à 95°C.

Les ARN sont chargés dans des puits de 0,5cm, sur gel 6% polyacrylamide (acrylamide/bisacrylamide 19 :1) / TBE 1X / 8,3M urée. La séparation des ARN est effectuée à 1000V constant, pendant 45 minutes. La position des ARN sur le gel est analysée par Moluclar Imager Pharose FX (BioRad).

Composition milieu réactionnel	Volume final = 20 μ l
Polyéthylène glycol PEG 8000 8%	5 μ l
EDTA ou MgAc 100mM	0,4 μ l
KAc 1M	1,5 μ l
ATP 100mM	0,4 μ l
Créatine phosphate 1M	0,4 μ l

Créatine kinase 10mg/mL	0,2µl
Nonidet P-40 0,5%	0,4µl
Dithiotréitol 0,1M	0,2µl
BSA (Bovin Serum Albumine) 10mg/mL	0,2µl
RNAguard (inhibiteur de RNase) 34U/µL	0,1µl
Extrait protéique	2-4µl
Substrat ARN marqué (20-40 fmoles de CYC1 ou CYC1-pré) repris dans l'eau	Qsp 20µl

Composition tampon Protéinase K	Volumes
Tampon 2X (200mM Tris-HCL pH 7.9, 300mM NaCl, 25mM EDTA, 2% SDS)	50µl
Protéinase K (20mg/mL)	1µl
Glycogène (20mg/mL)	0,25µl
Eau	23,75µl

5.3. Biochimie

5.3.1. Purification des protéines

Les culots de bactérie sont dilués dans du Tris pH 7,5, 25 mM et lysés. La lyse peut être effectuée par emulsifier, auquel cas environ trois passage à pression 1000 psi seront nécessaire, ou par sonication, et dans ce cas nous soniquerons trois fois trentes secondes à 4 W. Dans les deux cas l'échantillon restera dans la glace dans la mesure du possible. La fraction

insoluble est alors culotée (rotor Avanti 25.50, 50 000 g, 1h, 4°C) et le surnageant est incubé une heure avec la résine à 4°C (HiTrap Ni²⁺, Sigma). La résine est alors récupérée et tassée dans un adaptateur (XK16, Amersham Biosciences) installé sur un Akta Purifier. Le protocole suivi varie alors selon le matériel à purifier.

Les étapes de concentrations sont effectuées à l'aide de concentrateur Millipore.

5.3.2. Purification des ARN

Les ARNs issus de transcription *in vitro* ou de synthèse, commandés au service interne à l'IECB. Selon la procédure de synthèse et l'opérateur, nous pouvons les recevoir soit encore protégés et liés à la matrice de synthèse soit prêts à l'emploi. La méthodologie générale consiste à décrocher l'extrémité 5' de la matrice, déprotéger l'extrémité 3'OH et finalement purifier sur gel. Le protocole sera adapté au cas par cas, selon l'état de l'ARN que nous recevons.

La séparation de la matrice est effectuée par une incubation des billes liés aux ARNs dans 1ml de NH₄OH/EtOH (1 :4) pendant 16h 55°C ou 2h à 80°C. Après avoir refroidit, la réaction est centrifugée et le surnageant est récupéré. Les billes sont rincées trois fois avec 250 µl EtOH/Acétonitrile/H₂O (3 :1 :1) en récupérant la solution de lavage à chaque étape. Les ARNs sont alors récupérés par évaporation de la solution au SpeedVac.

La déprotection de l'extrémité 3'OH est réalisée en resuspendant les ARNs dans 250 µl de solution de déprotection (N-méthylpyrrolidinone/TEA/TEA-3HF [9 :5 :6]) et en laissant incubé 1h30 à 65°C, en homogénéisant régulièrement. 1 ml de butanol est ajouté au mélange refroidit, et le tout est centrifugé 1 min à 14500 rpm pour pouvoir ôter le surnageant. Le culot est repris dans 200 µl d'eau sans RNase et 1.8 ml de butanol sont ajoutés. Le culot est resuspendu par vortex puis centrifugé 1 min à 14500 rpm pour ôter le surnageant. Le culot d'ARN est séché au SpeedVac puis repris dans de l'eau non contaminée.

L'échantillon est alors purifié sur un gel d'acrylamide 12% (19 :1) 7M Urée en TBE. Le gel est pré-migré 20 min à 35 W puis l'échantillon est chargé pour une migration à 40 W, environ 4 h. Le gel est alors révélé par une illumination à 254 nm et la bande d'intérêt est excisée. L'ARN est alors électroélué dans du TBE et précipité sur la nuit à -20°C par l'ajout de 1 volume d'EtOH 100% et 0.1 volume de NaAc 1M pH 5.3. L'ARN est alors centrifugé 2h30 à

3220g, 4°C et le culot est repris dans 35 ml EtOH 70% L'échantillon est alors centrifugé 40 min à 15000g, 4°C et le culot est repris dans 500 µl d'eau sans RNAses et la concentration est mesurée au Nanodrop.

5.3.3. Cristallogenèse

Pour obtenir des cristaux de nos protéines intérêt, les échantillons sont concentrés entre 3 à 20 mg/ml et des essais de cristallogenèse en goutte assise sont effectués. Les gouttes sont réalisées avec une très bonne reproductibilité grâce au robot de cristallisation Mosquito et aux kits de conditions commerciaux disponibles au laboratoire. Les gouttes seront généralement constituées par 250 nl d'échantillon et 250 nl de condition de cristallisation et sont réalisées dans des boîtes 96 puits Grenier. La boîte scellée est ensuite stockée dans une armoire thermostatée (Molecular Dimensions) à humidité contrôlée pour que la diffusion de vapeur se fasse dans les conditions les plus stables possibles. La stratégie expérimentale classique pour une nouvelle construction consistera à effectuer un panel de quatre à six boîtes avec des kits généraux, comme le PACT, l'Index, le JCSG+ou le Classics, avec un échantillon autour de 10 mg/ml. Les boîtes seront ensuite observées au microscope à la recherche de conditions ayant permis la nucléation. Les conditions identifiées par cette stratégie nous permettront alors de concevoir des plaques explorant plus finement l'importance des différents composants de la condition initiale. Une fois la condition initiale optimisée, l'utilisation du kit Additifs permettra une exploration rapide de potentiels agents améliorant les cristaux.

Kits disponibles au laboratoire :

Nom	Fournisseur
PACT suite	Qiagen
JCSG+ suite	Qiagen
NeXtal tubes	Qiagen
Cryos suite	Qiagen
Classics suite	Qiagen
PEGs II suite	Qiagen
MbClass suite	Qiagen

MbClass II suite	Qiagen
ProComplex suite	Qiagen
PEGs suite	Qiagen
Anions suite	Qiagen
Cations suite	Qiagen
Midas I	Molecular Dimensions
Midas il	Molecular Dimensions
Proplex I	Molecular Dimensions
Ploplex II	Molecular Dimensions
Morpheus	Molecular Dimensions
Index	Hampton Research
Additives screen	Hampton Research

5.3.4. GraFix

La préparation des échantillons en GraFix est effectuée comme décrite dans la publication établissant la technique (Kastner, Fischer et al. 2008).

Un gradient de 10 à 30% de glycérol et 0 à 0,2% de glutaraldéhyde est préparé (Gradient Master plateau) dans des tubes de polyallomer 11*60mm et 20 à 66 nmol, pour un volume maximal de 200µl, de complexe sont chargés sur le sommet du gradient. Pour chaque essai de crosslink, au moins un tube témoin sera effectué sans agent crosslinkant. Les échantillons sont alors centrifugés 15h à 50000rpm à 4°C (rotor SW60Ti, centrifugeuse OPTIMA L-80XP, BECKMAN-COULTER) puis les tubes sont montés sur un système maison où une pompe péristaltique couplée à un collecteur de fractions va permettre d'injecter une solution à 60% glycérol au fond du tube, sans perturber le gradient. Le collecteur de fraction couplé à une cellule UV permettra de détecter les fractions contenant le matériel. Mais les fractions issues de gradients sans agent crosslinquant seront précipité avec du TCA et migrées sur un gel SDS-PAGE pour ne pas laisser de doutes sur les fractions contenant les protéines. Une fois identifiées, les fraction contenant le facteur ponté seront congelée dans l'azote liquide et envoyée à Patrick Bron.

5.3.5. Crosslink des protéines pour cristallogénèse

Etant donné que je rencontre visiblement un problème de stabilité et/ou de dégradation de CF IA durant la cristallogénèse, une des pistes évoquées pour y remédier fut le pontage du facteur avant les essais. Une recherche bibliographique ne me permit pas de trouver un protocole ou une structure résolue par cette approche. Par contre, il est beaucoup plus classique de stabiliser des cristaux déjà formés grâce à un agent pontant à faible concentration via à diffusion de vapeur. Le plus classiquement utilisé est le glutaraldehyde et je me suis également tourné vers lui.

Pour favoriser le pontage intra-CF IA, j'ai travaillé avec mes fractions de sortie de colonne d'exclusion de taille, soit environ 8 ml. Ce volume étant trop important pour pouvoir essayer le protocole par diffusion de vapeur, j'ai directement ajouté 0.05% de glutaraldehyde dans mon échantillon à 4°C. Après une homogénéisation douce à la pipette, et une incubation de 5 min, j'ai stoppé la réaction de pontage par ajout de Tris 1M pH 7.5, pour une concentration finale de 50mM. J'ai une fois de plus homogénéisé délicatement et laissé incuber 1h30 à 4°C pour avoir une désactivation totale du glutaraldehyde. Une fois ceci fait, j'ai concentré l'échantillon par centrifugation dans des Amicon Ultra (Millipore) possédant un seuil de filtration à 30 000 Da.

5.4. Biophysique

5.4.1. SAXS

Les acquisitions de données SAXS ont été faites sur la ligne de lumière SWING au synchrotron SOLEIL ($\lambda = 1.033 \text{ \AA}$). Le détecteur (PCCD170170 [AVIEX]) est positionné pour acquérir des données dans une gamme de $0,008 > q > 0,33 \text{ \AA}^{-1}$ ($q = 4\pi \sin\theta \lambda^{-1}$, où 2θ est l'angle de diffusion).

Tous les échantillons sont centrifugés 10 min à 14000g avant toute analyse afin de précipiter de potentiels agrégats. La concentration est alors vérifiée par absorption UV à $\lambda = 280 \text{ nm}$ sur un spectrophotomètre NanoDrop 1000 (Thermo Scientific). Pour tous les échantillons, la concentration est portée à 10 mg/ml, et quand la quantité de matériel disponible est suffisante, deux échantillons sont également préparés à 2 et 4 mg/ml. 50 μl

d'échantillon sont injectés dans une colonne d'exclusion sphérique (Agilent Bio SEC-3 HPLC column, 300 Å) et élués directement dans le capillaire d'acquisition du SAXS. Le tampon d'élution est du Tris pH 7.5 25 mM, NaCl 250 mM, DTT 1 mM. Des données de SAXS sont acquises durant l'élution du pic et durant les premières minutes de l'élution, ces fenêtres d'acquisition servant de référence pour la diffusion du tampon. Le traitement des données de diffusion, la moyenne des acquisitions ainsi que la soustraction de la contribution du solvant sont réalisés avec FOXTROT. Les traitements subséquents, l'analyse et les modélisations sont réalisés avec des programmes de la suite ATSAS.

5.4.2. Cristallographie aux rayons X

Les cristaux sont pêchés à l'aide de boucles en nylon dans les gouttes de cristallogenèse et congelé en « flash-freeze » dans de l'azote liquide, après un passage dans une solution de cryogénisation selon les cas.

Les cristaux sont ensuite testé en diffraction sur la ligne de lumière PX (synchrotron SOLEIL) ou ID29 (synchrotron ESRF). Les jeux de données sont traités avec XDS, le remplacement moléculaire est effectué avec Phenix et les affinements avec Buster. Les étapes de constructions sont alors effectuées dans Coot.

5.4.3. Microscopie électronique

Les complexes sont déposés sur une grille de cuivre couverte de cuivre pendant deux minutes. L'excès d'échantillon est retiré par capillarité et le colorant est ajouté pendant une minute. Les grilles sont alors séchées et observées sur un microscope JEOL 2200 FS à la recherche d'objets. Pour les essais de cryo-microscopie, la grille est plongée dans l'azote liquide immédiatement après le dépôt de l'échantillon, puis observée sous un flux d'azote.

5.4.4. Spectrométrie de masse

Avant toute analyse, le tampon de la protéine est échangé contre une solution d'acétate d'ammonium à 250mM, pH 7,5 par ultracentrifugation (Vivaspin 500 MWCO 10000 to 30000, Sartorius, Göttingen, DE) ou micro gel filtration (Zeba 0.5 ml, Thermo Scientific, Rockford, IL). Les essais de nanoESI-MA sur le CF IA entier sont effectués sur un spectromètre de masse electrospray Q-TOF (Synapt G1, Waters, Manchester, UK), en utilisant des aiguilles

de quartz (QuartzTip, New Objective, Woburn, USA). Les sous-complexes ont été caractérisés sur un électrospray Q-TOF (Synapt G2 HDMS, Waters, Manchester, UK) et un electrospray TOF (LCT, Waters, Manchester, UK) équipé par une source nanoESI (Triversa Nanomate, Advion Biosciences, Ithaca, NY) en mode ion positif.

Les analyses en conditions dénaturantes sont effectuées en injectant 2 μ M de protéine dans un tampon 1:1 (v/v) eau/acétonitrile avec 1% (v/v) acide formique. Les analyses en conditions natives sont réalisées avec 5 μ M de protéines dans du tampon NH_4Ac , 250 mM, pH 7,5.

Les analyses en LC-MS ont été réalisées en injectant 0,1 à 1 nmol de protéine dans une colonne HPLC phase reverse C8 Vydac (250mm x 2.1mm, Grace, Columbia, USA). Avec un système HPLC (Agilent LC-1100 series, Agilent, Waldbronn, DE) couplé à un spectromètre Q-TOF (MicroTOF, Bruker, Bremen, DE). L'élution est alors réalisé avec un gradient de 30 à 90% de tampon B (CH_3CN + 0.08% TFA; solvant A: H_2O + 0.1% TFA) à 200 μ l/min.

5.4.5. Absorption atomique

Les échantillons purifiés sont concentrés jusqu'à 3-10 μ M pour un volume de 500-100 μ l et congelés en azote liquide. Les mesures sont faites sur un spectromètre d'absorption atomique Varian AA220. 100 μ l d'échantillon sont injectés dans un brûleur air/acétylène et le maximum d'absorption à 213,9 nm est enregistré. Une gamme de ZnSO_4 (1,25 à 20 μ M) faite dans le tampon des échantillons permet de calculer la concentration de Zn^{2+} présente dans les échantillons.

5.4.6. SPR

Les analyses de résonance plasmonique de surface sont réalisées avec un appareil Biacore T200 (Biacore) et des puces de détection CM5 (Biacore). La surface a été activée selon les instructions du constructeur avec le tampon NHS + EDC (Biacore). 100 à 400 RU de protéine sont immobilisés sur une piste de la puce en injectant les échantillons dilués dans du NaCl 100 mM, Phosphate 10 mM, pH 7,2. L'ARN est dilué dans le même tampon est injecté sur la protéine immobilisé à 20 μ l/min. Les cinétiques d'interactions sont calculées avec la méthode single cell kinetics (SCK) dans laquelle l'échantillon est injecté séquentiellement par

concentration croissante et l'étape de régénération est effectuée à la fin du cycle. Les données sont ensuite analysées en cherchant l'équation de la courbe selon un modèle de Langmuir 1:1.

6. Bibliographie

- Adams, P. D., P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart (2010). "PHENIX: a comprehensive Python-based system for macromolecular structure solution." Acta Crystallographica Section D **66**(2): 213-221.
- Ahn, S. H., M. Kim and S. Buratowski (2004). "Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing." Mol Cell **13**(1): 67-76.
- Al Husini, N., P. Kudla and A. Ansari (2013). "A Role for CF1A 3' End Processing Complex in Promoter-Associated Transcription." PLoS Genet **9**(8): e1003722.
- Amrani, N., M. Minet, M. Le Gouar, F. Lacroute and F. Wyers (1997). "Yeast Pab1 interacts with Rna15 and participates in the control of the poly(A) tail length in vitro." Mol Cell Biol **17**(7): 3694-3701.
- Amrani, N., M. Minet, F. Wyers, M. E. Dufour, L. P. Aggerbeck and F. Lacroute (1997). "PCF11 encodes a third protein component of yeast cleavage and polyadenylation factor I." Mol Cell Biol **17**(3): 1102-1109.
- Ansari, A. and M. Hampsey (2005). "A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping." Genes Dev **19**(24): 2969-2978.
- Bai, Y., T. C. Auperin, C. Y. Chou, G. G. Chang, J. L. Manley and L. Tong (2007). "Crystal Structure of Murine CstF-77: Dimeric Association and Implications for Polyadenylation of mRNA Precursors." Mol. Cell **25**(6): 863-875.
- Bai, Y., T. C. Auperin and L. Tong (2007). "The use of in situ proteolysis in the crystallization of murine CstF-77." Acta Crystallogr Sect F Struct Biol Cryst Commun **63**(Pt 2): 135-138.
- Barabino, S. M., W. Hubner, A. Jenny, L. Minvielle-Sebastia and W. Keller (1997). "The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins." Genes Dev **11**(13): 1703-1716.
- Bard, J., A. M. Zhelkovsky, S. Helmling, T. N. Earnest, C. L. Moore and A. Bohm (2000). "Structure of yeast poly(A) polymerase alone and in complex with 3'-dATP." Science **289**(5483): 1346-1349.
- Barnwal, R. P., S. D. Lee, C. Moore and G. Varani (2012). "Structural and biochemical analysis of the assembly and function of the yeast pre-mRNA 3' end processing complex CF I." Proc Natl Acad Sci U S A.
- Beaudoing, E., S. Freier, J. R. Wyatt, J. M. Claverie and D. Gautheret (2000). "Patterns of variant polyadenylation signal usage in human genes." Genome Res **10**(7): 1001-1010.
- Beelman, C. A. and R. Parker (1995). "Degradation of mRNA in eukaryotes." Cell **81**(2): 179-183.
- Bird, G., N. Fong, J. C. Gatlin, S. Farabaugh and D. L. Bentley (2005). "Ribozyme cleavage reveals connections between mRNA release from the site of transcription and pre-mRNA processing." Mol Cell **20**(5): 747-758.
- Birse, C. E., L. Minvielle-Sebastia, B. A. Lee, W. Keller and N. J. Proudfoot (1998). "Coupling termination of transcription to messenger RNA maturation in yeast." Science **280**(5361): 298-301.

Bonneaud, N., L. Minvielle-Sebastia, C. Cullin and F. Lacroute (1994). "Cellular localization of RNA14p and RNA15p, two yeast proteins involved in mRNA stability." J Cell Sci **107 (Pt 4)**: 913-921.

Buratowski, S. (2005). "Connections between mRNA 3' end processing and transcription termination." Curr Opin Cell Biol **17**(3): 257-261.

Buratowski, S. (2009). "Progression through the RNA polymerase II CTD cycle." Mol Cell **36**(4): 541-546.

Butler, J. S. and T. Platt (1988). "RNA processing generates the mature 3' end of yeast CYC1 messenger RNA in vitro." Science **242**(4883): 1270-1274.

Butler, J. S., P. P. Sadhale and T. Platt (1990). "RNA processing in vitro produces mature 3' ends of a variety of *Saccharomyces cerevisiae* mRNAs." Mol Cell Biol **10**(6): 2599-2605.

Callebaut, I., D. Moshous, J. P. Mornon and J. P. de Villartay (2002). "Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family." Nucleic Acids Res **30**(16): 3592-3601.

Calvo, O. and J. L. Manley (2005). "The transcriptional coactivator PC4/Sub1 has multiple functions in RNA polymerase II transcription." EMBO J **24**(5): 1009-1020.

Chen-Kiang, S., D. J. Wolgemuth, M. T. Hsu and J. E. Darnell, Jr. (1982). "Transcription and accurate polyadenylation in vitro of RNA from the major late adenovirus 2 transcription unit." Cell **28**(3): 575-584.

Chen, F., C. C. MacDonald and J. Wilusz (1995). "Cleavage site determinants in the mammalian polyadenylation signal." Nucleic Acids Res **23**(14): 2614-2620.

Chen, J. and C. Moore (1992). "Separation of factors required for cleavage and polyadenylation of yeast pre-mRNA." Mol Cell Biol **12**(8): 3470-3481.

Chen, Z., Y. Li and R. M. Krug (1999). "Influenza A virus NS1 protein targets poly(A)-binding protein II of the cellular 3'-end processing machinery." Embo J **18**(8): 2273-2283.

Cheng, H., X. He and C. Moore (2004). "The essential WD repeat protein Swd2 has dual functions in RNA polymerase II transcription termination and lysine 4 methylation of histone H3." Mol Cell Biol **24**(7): 2932-2943.

Cho, E. J., M. S. Kobor, M. Kim, J. Greenblatt and S. Buratowski (2001). "Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain." Genes Dev **15**(24): 3319-3329.

Colgan, D. F. and J. L. Manley (1997). "Mechanism and regulation of mRNA polyadenylation." Genes Dev **11**(21): 2755-2766.

Connelly, S. and J. L. Manley (1988). "A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II." Genes Dev **2**(4): 440-452.

Cooke, C. and J. C. Alwine (1996). "The cap and the 3' splice site similarly affect polyadenylation efficiency." Mol Cell Biol **16**(6): 2579-2584.

Cooke, C., H. Hans and J. C. Alwine (1999). "Utilization of splicing elements and polyadenylation signal elements in the coupling of polyadenylation and last-intron removal." Mol Cell Biol **19**(7): 4971-4979.

Cooper, D. L. and W. F. Marzluff (1978). "Polyadenylation of RNA in a cell-free system from mouse myeloma cells." J Biol Chem **253**(23): 8375-8380.

Cramer, P., D. A. Bushnell, J. Fu, A. L. Gnatt, B. Maier-Davis, N. E. Thompson, R. R. Burgess, A. M. Edwards, P. R. David and R. D. Kornberg (2000). "Architecture of RNA polymerase II and implications for the transcription mechanism." Science **288**(5466): 640-649.

Cramer, P., D. A. Bushnell and R. D. Kornberg (2001). "Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution." Science **292**(5523): 1863-1876.

Cui, M., M. A. Allen, A. Larsen, M. Macmorris, M. Han and T. Blumenthal (2008). "Genes involved in pre-mRNA 3'-end formation and transcription termination revealed by a lin-15 operon Muv suppressor screen." Proc Natl Acad Sci U S A **105**(43): 16665-16670.

Danckwardt, S., M. W. Hentze and A. E. Kulozik (2008). "3' end mRNA processing: molecular mechanisms and implications for health and disease." Embo J **27**(3): 482-498.

Dantoni, J. C., K. G. Murthy, J. L. Manley and L. Tora (1997). "Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA." Nature **389**(6649): 399-402.

Darnell, J. E., R. Wall and R. J. Tushinski (1971). "An adenylic acid-rich sequence in messenger RNA of HeLa cells and its possible relationship to reiterated sites in DNA." Proc Natl Acad Sci U S A **68**(6): 1321-1325.

Das, K., L. C. Ma, R. Xiao, B. Radvansky, J. Aramini, L. Zhao, J. Marklund, R. L. Kuo, K. Y. Twu, E. Arnold, R. M. Krug and G. T. Montelione (2008). "Structural basis for suppression of a host antiviral response by influenza A virus." Proc Natl Acad Sci U S A **105**(35): 13093-13098.

Dheur, S., T. A. Voile, F. Voisin-Hakil, M. Minet, J. M. Schmitter, F. Lacroute, F. Wyers and L. Minvielle-Sebastia (2003). "Pti1p and Ref2p found in association with the mRNA 3' end formation complex direct snoRNA maturation." EMBO J **22**(11): 2831-2840.

Dichtl, B., D. Blank, M. Ohnacker, A. Friedlein, D. Roeder, H. Langen and W. Keller (2002). "A role for SSU72 in balancing RNA polymerase II transcription elongation and termination." Mol Cell **10**(5): 1139-1150.

Dichtl, B., D. Blank, M. Sadowski, W. Hubner, S. Weiser and W. Keller (2002). "Yhh1p/Cft1p directly links poly(A) site recognition and RNA polymerase II transcription termination." Embo J **21**(15): 4125-4135.

Dichtl, B. and W. Keller (2001). "Recognition of polyadenylation sites in yeast pre-mRNAs by cleavage and polyadenylation factor." Embo J **20**(12): 3197-3209.

Diebold, M. L., S. Fribourg, M. Koch, T. Metzger and C. Romier (2011). "Deciphering correct strategies for multiprotein complex assembly by co-expression: application to complexes as large as the histone octamer." J Struct Biol **175**(2): 178-188.

Dikfidan, A., B. Loll, C. Zeymer, I. Magler, T. Clausen and A. Meinhart (2014). "RNA Specificity and Regulation of Catalysis in the Eukaryotic Polynucleotide Kinase Clp1." Mol Cell.

Dominski, Z., A. J. Carpousis and B. Clouet-d'Orval (2013). "Emergence of the beta-CASP ribonucleases: highly conserved and ubiquitous metallo-enzymes involved in messenger RNA maturation and degradation." Biochim Biophys Acta **1829**(6-7): 532-551.

- Dominski, Z. and W. F. Marzluff (1999). "Formation of the 3' end of histone mRNA." Gene **239**(1): 1-14.
- Dominski, Z. and W. F. Marzluff (2007). "Formation of the 3' end of histone mRNA: getting closer to the end." Gene **396**(2): 373-390.
- Dosztanyi, Z., V. Csizmok, P. Tompa and I. Simon (2005). "IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content." Bioinformatics **21**(16): 3433-3434.
- Dupin, A. F. and S. Fribourg (2014). "Structural basis for ATP loss by Clp1p in a G135R mutant protein." Biochimie.
- Edmonds, M. and R. Abrams (1960). "Polynucleotide biosynthesis: formation of a sequence of adenylate units from adenosine triphosphate by an enzyme from thymus nuclei." J Biol Chem **235**: 1142-1149.
- Edmonds, M., M. H. Vaughan, Jr. and H. Nakazato (1971). "Polyadenylic acid sequences in the heterogeneous nuclear RNA and rapidly-labeled polyribosomal RNA of HeLa cells: possible evidence for a precursor relationship." Proc Natl Acad Sci U S A **68**(6): 1336-1340.
- Edwards-Gilbert, G., K. L. Veraldi and C. Milcarek (1997). "Alternative poly(A) site selection in complex transcription units: means to an end?" Nucleic Acids Res **25**(13): 2547-2561.
- Emsley, P., B. Lohkamp, W. G. Scott and K. Cowtan (2010). "Features and development of Coot." Acta Crystallogr D Biol Crystallogr **66**(Pt 4): 486-501.
- Evrard, G., F. Mareuil, F. Bontems, C. Sizun and J. Perez (2011). "DADIMODO: a program for refining the structure of multidomain proteins and complexes against small-angle scattering data and NMR-derived restraints." Journal of Applied Crystallography **44**(6): 1264-1271.
- Ezeokonkwo, C., M. A. Ghazy, A. Zhelkovsky, P.-C. Yeh and C. Moore (2012). "Novel interactions at the essential N-terminus of poly(A) polymerase that could regulate poly(A) addition in *Saccharomyces cerevisiae*." FEBS Letters **586**(8): 1173-1178.
- Ezeokonkwo, C., A. Zhelkovsky, R. Lee, A. Bohm and C. L. Moore (2011). "A flexible linker region in Fip1 is needed for efficient mRNA polyadenylation." RNA **17**(4): 652-664.
- Flaherty, S. M., P. Fortes, E. Izaurralde, I. W. Mattaj and G. M. Gilmartin (1997). "Participation of the nuclear cap binding complex in pre-mRNA 3' processing." Proc Natl Acad Sci U S A **94**(22): 11893-11898.
- Fong, N. and D. L. Bentley (2001). "Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD." Genes Dev **15**(14): 1783-1795.
- Franke, D. and D. I. Svergun (2009). "DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering." Journal of Applied Crystallography **42**(2): 342-346.
- Garas, M., B. Dichtl and W. Keller (2008). "The role of the putative 3' end processing endonuclease Ysh1p in mRNA and snoRNA synthesis." RNA **14**(12): 2671-2684.
- Garcia-Gimeno, M. A., I. Munoz, J. Arino and P. Sanz (2003). "Molecular characterization of Ypi1, a novel *Saccharomyces cerevisiae* type 1 protein phosphatase inhibitor." J Biol Chem **278**(48): 47744-47752.

Garrido-Lecca, A. and T. Blumenthal (2010). "RNA polymerase II C-terminal domain phosphorylation patterns in *Caenorhabditis elegans* operons, polycistronic gene clusters with only one promoter." Mol Cell Biol **30**(15): 3887-3893.

Gavin, A. C., M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J. M. Rick, A. M. Michon, C. M. Cruciat, M. Remor, C. Hofert, M. Schelder, M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M. A. Heurtier, R. R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes, M. Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer and G. Superti-Furga (2002). "Functional organization of the yeast proteome by systematic analysis of protein complexes." Nature **415**(6868): 141-147.

Ghazal, G., J. Gagnon, P. E. Jacques, J. R. Landry, F. Robert and S. A. Elela (2009). "Yeast RNase III triggers polyadenylation-independent transcription termination." Mol Cell **36**(1): 99-109.

Ghazy, M. A., J. M. Gordon, S. D. Lee, B. N. Singh, A. Bohm, M. Hampsey and C. Moore (2011). "The interaction of Pcf11 and Clp1 is needed for mRNA 3'-end formation and is modulated by amino acids in the ATP-binding site." Nucleic Acids Res.

Ghazy, M. A., J. M. Gordon, S. D. Lee, B. N. Singh, A. Bohm, M. Hampsey and C. Moore (2012). "The interaction of Pcf11 and Clp1 is needed for mRNA 3'-end formation and is modulated by amino acids in the ATP-binding site." Nucleic Acids Res **40**(3): 1214-1225.

Ghazy, M. A., X. He, B. N. Singh, M. Hampsey and C. Moore (2009). "The essential N terminus of the Pta1 scaffold protein is required for snoRNA transcription termination and Ssu72 function but is dispensable for pre-mRNA 3'-end processing." Mol Cell Biol **29**(8): 2296-2307.

Ghosh, A., S. Shuman and C. D. Lima (2008). "The structure of Fcp1, an essential RNA polymerase II CTD phosphatase." Mol Cell **32**(4): 478-490.

Gilmartin, G. M., E. S. Fleming, J. Oetjen and B. R. Graveley (1995). "CPSF recognition of an HIV-1 mRNA 3'-processing enhancer: multiple sequence contacts involved in poly(A) site definition." Genes Dev **9**(1): 72-83.

Glover-Cutter, K., S. Kim, J. Espinosa and D. L. Bentley (2008). "RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes." Nat Struct Mol Biol **15**(1): 71-78.

Gnatt, A. L., P. Cramer, J. Fu, D. A. Bushnell and R. D. Kornberg (2001). "Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution." Science **292**(5523): 1876-1882.

Gonatopoulos-Pournatzis, T. and V. H. Cowling (2014). "Cap-binding complex (CBC)." Biochem J **457**(2): 231-242.

Gordon, J. M., S. Shikov, J. N. Kuehner, M. Liriano, E. Lee, W. Stafford, M. B. Poulsen, C. Harrison, C. Moore and A. Bohm (2011). "Reconstitution of CF IA from Overexpressed Subunits Reveals Stoichiometry and Provides Insights into Molecular Topology." Biochemistry.

Goujon, M., H. McWilliam, W. Li, F. Valentin, S. Squizzato, J. Paern and R. Lopez (2010). "A new bioinformatics analysis tools framework at EMBL-EBI." Nucleic Acids Research **38**(suppl 2): W695-W699.

Graber, J. H. (2003). "Variations in yeast 3'-processing cis-elements correlate with transcript stability." Trends Genet **19**(9): 473-476.

Graber, J. H., C. R. Cantor, S. C. Mohr and T. F. Smith (1999). "Genomic detection of new yeast pre-mRNA 3'-end-processing signals." Nucleic Acids Res **27**(3): 888-894.

Graveley, B. R., E. S. Fleming and G. M. Gilmartin (1996). "RNA structure is a critical determinant of poly(A) site recognition by cleavage and polyadenylation specificity factor." Mol Cell Biol **16**(9): 4942-4951.

Gromak, N., S. West and N. J. Proudfoot (2006). "Pause sites promote transcriptional termination of mammalian RNA polymerase II." Mol Cell Biol **26**(10): 3986-3996.

Gross, S. and C. Moore (2001). "Five subunits are required for reconstitution of the cleavage and polyadenylation activities of *Saccharomyces cerevisiae* cleavage factor I." Proc. Natl. Acad. Sci. U. S. A. **98**(11): 6080-6085.

Gross, S. and C. L. Moore (2001). "Rna15 interaction with the A-rich yeast polyadenylation signal is an essential step in mRNA 3'-end formation." Mol. Cell. Biol. **21**(23): 8045-8055.

Gu, B., D. Eick and O. Bensaude (2013). "CTD serine-2 plays a critical role in splicing and termination factor recruitment to RNA polymerase II in vivo." Nucleic Acids Res **41**(3): 1591-1603.

Guo, Z. and F. Sherman (1996). "Signals sufficient for 3'-end formation of yeast mRNA." Mol Cell Biol **16**(6): 2772-2776.

Haddad, R. I., F. d. r. Maurice, N. Viphakone, F. Voisinet-Hakil, S. b. Fribourg and L. Minvielle-SV@bastia (2011). "An essential role for Clp1 in assembly of polyadenylation complex CF IA and Pol II transcription termination." Nucleic Acids Research.

Hans, H. and J. C. Alwine (2000). "Functionally Significant Secondary Structure of the Simian Virus 40 Late Polyadenylation Signal." Molecular and Cellular Biology **20**(8): 2926-2932.

He, X., A. U. Khan, H. Cheng, D. L. Pappas, Jr., M. Hampsey and C. L. Moore (2003). "Functional interactions between the transcription and mRNA 3' end processing machineries mediated by Ssu72 and Sub1." Genes Dev **17**(8): 1030-1042.

He, X. and C. Moore (2005). "Regulation of Yeast mRNA 3',5'-End Processing by Phosphorylation." Molecular Cell **19**(5): 619-629.

Helmling, S., A. Zhelkovsky and C. L. Moore (2001). "Fip1 regulates the activity of Poly(A) polymerase through multiple interactions." Mol Cell Biol **21**(6): 2026-2037.

Higgs, D. R., S. E. Goodbourn, J. Lamb, J. B. Clegg, D. J. Weatherall and N. J. Proudfoot (1983). "Alpha-thalassaemia caused by a polyadenylation signal mutation." Nature **306**(5941): 398-400.

Hirose, Y. and J. L. Manley (1998). "RNA polymerase II is an essential mRNA polyadenylation factor." Nature **395**(6697): 93-96.

Holbein, S., S. Scola, B. Loll, B. S. Dichtl, W. Hübner, A. Meinhart and B. Dichtl (2011). "The P-Loop Domain of Yeast Clp1 Mediates Interactions Between CF IA and CPF Factors in Pre-mRNA 3',5'-End Formation." PLoS ONE **6**(12): e29139.

Hollingworth, D., C. G. Noble, I. A. Taylor and A. Ramos (2006). "RNA polymerase II CTD phosphopeptides compete with RNA for the interaction with Pcf11." Rna **12**(4): 555-560.

Hu, J., C. S. Lutz, J. Wilusz and B. Tian (2005). "Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation." RNA **11**(10): 1485-1493.

Huez, G., G. Marbaix, D. Gallwitz, E. Weinberg, R. Devos, E. Hubert and Y. Cleuter (1978). "Functional stabilisation of HeLa cell histone messenger RNAs injected into *Xenopus* oocytes by 3'-OH polyadenylation." Nature **271**(5645): 572-573.

Huez, G., G. Marbaix, E. Hubert, Y. Cleuter, M. Leclercq, H. Chantrenne, R. Devos, H. Soreq, U. Nudel and U. Z. Littauer (1975). "Readenylation of polyadenylate-free globin messenger RNA restores its stability in vivo." Eur J Biochem **59**(2): 589-592.

Huez, G., G. Marbaix, E. Hubert, Y. Cleuter, M. Leclercq, H. Chantrenne, R. Devos, H. Soreq, R. Salomon, U. Nudel and U. Littauer (1975). "Proceedings: The polyadenylic segment of globin messenger RNA ensures its stability in vivo." Arch Int Physiol Biochim **83**(2): 370-371.

Jurica, M. S. and M. J. Moore (2003). "Pre-mRNA splicing: awash in a sea of proteins." Mol Cell **12**(1): 5-14.

Kabsch, W. (2010). "XDS." Acta Crystallogr D Biol Crystallogr **66**(Pt 2): 125-132.

Karaca, E., S. Weitzer, D. Pehlivan, H. Shiraishi, T. Gogakos, T. Hanada, S. N. Jhangiani, W. Wiszniewski, M. Withers, I. M. Campbell, S. Erdin, S. Isikay, L. M. Franco, C. Gonzaga-Jauregui, T. Gambin, V. Gelowani, J. V. Hunter, G. Yesil, E. Koparir, S. Yilmaz, M. Brown, D. Briskin, M. Hafner, P. Morozov, T. A. Farazi, C. Bernreuther, M. Glatzel, S. Trattnig, J. Friske, C. Kronnerwetter, M. N. Bainbridge, A. Gezdirici, M. Seven, D. M. Muzny, E. Boerwinkle, M. Ozen, T. Clausen, T. Tuschl, A. Yuksel, A. Hess, R. A. Gibbs, J. Martinez, J. M. Penninger and J. R. Lupski (2014). "Human CLP1 Mutations Alter tRNA Biogenesis, Affecting Both Peripheral and Central Nervous System Function." Cell **157**(3): 636-650.

Kastner, B., N. Fischer, M. M. Golas, B. Sander, P. Dube, D. Boehringer, K. Hartmuth, J. Deckert, F. Hauer, E. Wolf, H. Uchtenhagen, H. Urlaub, F. Herzog, J. M. Peters, D. Poerschke, R. Luhrmann and H. Stark (2008). "GraFix: sample preparation for single-particle electron cryomicroscopy." Nat Methods **5**(1): 53-55.

Keller, W., S. Bienroth, K. M. Lang and G. Christofori (1991). "Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3' processing signal AAUAAA." Embo J **10**(13): 4241-4249.

Keller, W. and L. Minvielle-Sebastia (1997). "A comparison of mammalian and yeast pre-mRNA 3'-end processing." Curr Opin Cell Biol **9**(3): 329-336.

Kennedy, S. A., M. L. Frazier, M. Steiniger, A. M. Mast, W. F. Marzluff and M. R. Redinbo (2009). "Crystal structure of the HEAT domain from the Pre-mRNA processing factor Symplekin." J Mol Biol **392**(1): 115-128.

Kessler, M. M., M. F. Henry, E. Shen, J. Zhao, S. Gross, P. A. Silver and C. L. Moore (1997). "Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3'-end formation in yeast." Genes Dev **11**(19): 2545-2556.

Kessler, M. M., J. Zhao and C. L. Moore (1996). "Purification of the *Saccharomyces cerevisiae* cleavage/polyadenylation factor I. Separation into two components that are required for both cleavage and polyadenylation of mRNA 3' ends." J Biol Chem **271**(43): 27167-27175.

Kim, M., N. J. Krogan, L. Vasiljeva, O. J. Rando, E. Nedea, J. F. Greenblatt and S. Buratowski (2004). "The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II." Nature **432**(7016): 517-522.

Kim, M., L. Vasiljeva, O. J. Rando, A. Zhelkovsky, C. Moore and S. Buratowski (2006). "Distinct pathways for snoRNA and mRNA termination." Mol Cell **24**(5): 723-734.

Kolev, N. G., T. A. Yario, E. Benson and J. A. Steitz (2008). "Conserved motifs in both CPSF73 and CPSF100 are required to assemble the active endonuclease for histone mRNA 3'-end maturation." EMBO Rep **9**(10): 1013-1018.

Komarnitsky, P., E. J. Cho and S. Buratowski (2000). "Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription." Genes Dev **14**(19): 2452-2460.

Konarev, P. V., V. V. Volkov, A. V. Sokolova, M. H. J. Koch and D. I. Svergun (2003). "PRIMUS: a Windows PC-based system for small-angle scattering data analysis." Journal of Applied Crystallography **36**(5): 1277-1282.

Krishnamurthy, S., X. He, M. Reyes-Reyes, C. Moore and M. Hampsey (2004). "Ssu72 Is an RNA polymerase II CTD phosphatase." Mol Cell **14**(3): 387-394.

Krogan, N. J., G. Cagney, H. Yu, G. Zhong, X. Guo, A. Ignatchenko, J. Li, S. Pu, N. Datta, A. P. Tikuisis, T. Punna, J. M. Peregrin-Alvarez, M. Shales, X. Zhang, M. Davey, M. D. Robinson, A. Paccanaro, J. E. Bray, A. Sheung, B. Beattie, D. P. Richards, V. Canadien, A. Lalev, F. Mena, P. Wong, A. Starostine, M. M. Canete, J. Vlasblom, S. Wu, C. Orsi, S. R. Collins, S. Chandran, R. Haw, J. J. Rilstone, K. Gandi, N. J. Thompson, G. Musso, P. St Onge, S. Ghanny, M. H. Lam, G. Butland, A. M. Altaf-Ul, S. Kanaya, A. Shilatifard, E. O'Shea, J. S. Weissman, C. J. Ingles, T. R. Hughes, J. Parkinson, M. Gerstein, S. J. Wodak, A. Emili and J. F. Greenblatt (2006). "Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*." Nature **440**(7084): 637-643.

Kyburz, A., A. Friedlein, H. Langen and W. Keller (2006). "Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3' end processing and splicing." Mol Cell **23**(2): 195-205.

Kyburz, A., M. Sadowski, B. Dichtl and W. Keller (2003). "The role of the yeast cleavage and polyadenylation factor subunit Ydh1p/Cft2p in pre-mRNA 3'-end formation." Nucleic Acids Res **31**(14): 3936-3945.

Lamanna, A. C. and K. Karbstein (2009). "Nob1 binds the single-stranded cleavage site D at the 3'-end of 18S rRNA with its PIN domain." Proc Natl Acad Sci U S A **106**(34): 14259-14264.

Lee, J. Y., Z. Ji and B. Tian (2008). "Phylogenetic analysis of mRNA polyadenylation sites reveals a role of transposable elements in evolution of the 3'-end of genes." Nucleic Acids Res **36**(17): 5581-5590.

Lee, S. Y., J. Mendecki and G. Brawerman (1971). "A polynucleotide segment rich in adenylic acid in the rapidly-labeled polyribosomal RNA component of mouse sarcoma 180 ascites cells." Proc Natl Acad Sci U S A **68**(6): 1331-1335.

Leeper, T. C., X. Qu, C. Lu, C. Moore and G. Varani (2010). "Novel Protein-Protein Contacts Facilitate mRNA 3'-Processing Signal Recognition by Rna15 and Hrp1." Journal of Molecular Biology **401**(3): 334-349.

Legendre, M. and D. Gautheret (2003). "Sequence determinants in human polyadenylation site selection." BMC Genomics **4**(1): 7.

Legrand, P., N. Pinaud, L. Minvielle-Sebastia and S. Fribourg (2007). "The structure of the CstF-77 homodimer provides insights into CstF assembly." Nucleic Acids Res **35**(13): 4515-4522.

Lenssen, E., N. James, I. Pedruzzi, F. Dubouloz, E. Cameroni, R. Bisig, L. Maillet, M. Werner, J. Roosen, K. Petrovic, J. Winderickx, M. A. Collart and C. De Virgilio (2005). "The Ccr4-Not complex independently controls both Msn2-dependent transcriptional activation--via a newly identified Glc7/Bud14 type I protein phosphatase module--and TFIID promoter distribution." Mol Cell Biol **25**(1): 488-498.

Li, T., X. Chen, K. C. Garbutt, P. Zhou and N. Zheng (2006). "Structure of DDB1 in complex with a paramyxovirus V protein: viral hijack of a propeller cluster in ubiquitin ligase." Cell **124**(1): 105-117.

Licatalosi, D. D., G. Geiger, M. Minet, S. Schroeder, K. Cilli, J. B. McNeil and D. L. Bentley (2002). "Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II." Mol. Cell **9**(5): 1101-1111.

Linding, R., R. B. Russell, V. Neduva and T. J. Gibson (2003). "GlobPlot: exploring protein sequences for globularity and disorder." Nucleic Acids Research **31**(13): 3701-3708.

Lindstrom, D. L., S. L. Squazzo, N. Muster, T. A. Burckin, K. C. Wachter, C. A. Emigh, J. A. McCleery, J. R. Yates, 3rd and G. A. Hartzog (2003). "Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins." Mol Cell Biol **23**(4): 1368-1378.

Logan, J., E. Falck-Pedersen, J. E. Darnell, Jr. and T. Shenk (1987). "A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene." Proc Natl Acad Sci U S A **84**(23): 8306-8310.

Loke, J. C., E. A. Stahlberg, D. G. Strenski, B. J. Haas, P. C. Wood and Q. Q. Li (2005). "Compilation of mRNA polyadenylation signals in Arabidopsis revealed a new signal element and potential secondary structures." Plant Physiol **138**(3): 1457-1468.

Lopez, F., S. Granjeaud, T. Ara, B. Ghattas and D. Gautheret (2006). "The disparate nature of "intergenic" polyadenylation sites." RNA **12**(10): 1794-1801.

Luo, W., A. W. Johnson and D. L. Bentley (2006). "The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model." Genes Dev **20**(8): 954-965.

Lutz, C. S., K. G. Murthy, N. Schek, J. P. O'Connor, J. L. Manley and J. C. Alwine (1996). "Interaction between the U1 snRNP-A protein and the 160-kD subunit of cleavage-polyadenylation specificity factor increases polyadenylation efficiency in vitro." Genes Dev **10**(3): 325-337.

Mandart, E. and R. Parker (1995). "Effects of mutations in the *Saccharomyces cerevisiae* RNA14, RNA15, and PAP1 genes on polyadenylation in vivo." Mol Cell Biol **15**(12): 6979-6986.

Mandel, C. R., Y. Bai and L. Tong (2008). "Protein factors in pre-mRNA 3'-end processing." Cell Mol Life Sci **65**(7-8): 1099-1122.

Mandel, C. R., S. Kaneko, H. Zhang, D. Gebauer, V. Vethantham, J. L. Manley and L. Tong (2006). "Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease." Nature **444**(7121): 953-956.

- Mangus, D. A., M. M. Smith, J. M. McSweeney and A. Jacobson (2004). "Identification of factors regulating poly(A) tail synthesis and maturation." Mol Cell Biol **24**(10): 4196-4206.
- Manley, J. L. (1983). "Accurate and specific polyadenylation of mRNA precursors in a soluble whole-cell lysate." Cell **33**(2): 595-605.
- Martin, G. and W. Keller (1996). "Mutational analysis of mammalian poly(A) polymerase identifies a region for primer binding and catalytic domain, homologous to the family X polymerases, and to other nucleotidyltransferases." EMBO J **15**(10): 2593-2603.
- Martin, G., W. Keller and S. Doublié (2000). "Crystal structure of mammalian poly(A) polymerase in complex with an analog of ATP." Embo J. **19**(16): 4193-4203.
- Mayer, A., M. Heidemann, M. Lidschreiber, A. Schrieck, M. Sun, C. Hintermair, E. Kremmer, D. Eick and P. Cramer (2012). "CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II." Science **336**(6089): 1723-1725.
- Mayer, A., A. Schrieck, M. Lidschreiber, K. Leike, D. E. Martin and P. Cramer (2012). "The spt5 C-terminal region recruits yeast 3' RNA cleavage factor I." Mol Cell Biol **32**(7): 1321-1331.
- McCoy, A. J. (2007). "Solving structures of protein complexes by molecular replacement with Phaser." Acta Crystallogr D Biol Crystallogr **63**(Pt 1): 32-41.
- McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read (2007). "Phaser crystallographic software." Journal of Applied Crystallography **40**(4): 658-674.
- McCracken, S., N. Fong, K. Yankulov, S. Ballantyne, G. Pan, J. Greenblatt, S. D. Patterson, M. Wickens and D. L. Bentley (1997). "The C-terminal domain of RNA polymerase II couples mRNA processing to transcription." Nature **385**(6614): 357-361.
- McWilliam, H., W. Li, M. Uludag, S. Squizzato, Y. M. Park, N. Buso, A. P. Cowley and R. Lopez (2013). "Analysis Tool Web Services from the EMBL-EBI." Nucleic Acids Research **41**(W1): W597-W600.
- Meinhart, A. and P. Cramer (2004). "Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors." Nature **430**(6996): 223-226.
- Meinke, G., C. Ezeokonkwo, P. Balbo, W. Stafford, C. Moore and A. Böhm (2008). "Structure of yeast poly(A) polymerase in complex with a peptide from Fip1, an intrinsically disordered protein." Biochemistry **47**(26): 6859-6869.
- Michalova, E., B. Vojtesek and R. Hrstka (2013). "Impaired pre-mRNA processing and altered architecture of 3' untranslated regions contribute to the development of human disorders." Int J Mol Sci **14**(8): 15681-15694.
- Millevoi, S. and S. Vagner (2010). "Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation." Nucleic Acids Res **38**(9): 2757-2774.
- Minvielle-Sebastia, L., P. J. Preker and W. Keller (1994). "RNA14 and RNA15 proteins as components of a yeast pre-mRNA 3'-end processing factor." Science **266**(5191): 1702-1705.
- Minvielle-Sebastia, L., P. J. Preker, T. Wiederkehr, Y. Strahm and W. Keller (1997). "The major yeast poly(A)-binding protein is associated with cleavage factor IA and functions in premessenger RNA 3'-end formation." Proc. Natl. Acad. Sci. U. S. A. **94**(15): 7897-7902.

Minvielle-Sebastia, L., B. Winsor, N. Bonneaud and F. Lacroute (1991). "Mutations in the yeast RNA14 and RNA15 genes result in an abnormal mRNA decay rate; sequence analysis reveals an RNA-binding domain in the RNA15 protein." Mol. Cell. Biol. **11**(6): 3075-3087.

Moore, C. L. and P. A. Sharp (1985). "Accurate cleavage and polyadenylation of exogenous RNA substrate." Cell **41**(3): 845-855.

Moore, M. J. and N. J. Proudfoot (2009). "Pre-mRNA processing reaches back to transcription and ahead to translation." Cell **136**(4): 688-700.

Moqtaderi, Z., J. V. Geisberg, Y. Jin, X. Fan and K. Struhl (2013). "Species-specific factors mediate extensive heterogeneity of mRNA 3' ends in yeasts." Proc Natl Acad Sci U S A.

Moreira, A., Y. Takagaki, S. Brackenridge, M. Wollerton, J. L. Manley and N. J. Proudfoot (1998). "The upstream sequence element of the C2 complement poly(A) signal activates mRNA 3' end formation by two distinct mechanisms." Genes Dev **12**(16): 2522-2534.

Moreno-Morcillo, M., L. Minvielle-Sebastia, S. Fribourg and C. D. Mackereth (2011). "Locked tether formation by cooperative folding of Rna14p monkeytail and Rna15p hinge domains in the yeast CF IA complex." Structure **19**(4): 534-545.

Moreno-Morcillo, M., L. Minvielle-Sebastia, C. Mackereth and S. Fribourg (2011). "Hexameric architecture of CstF supported by CstF-50 homodimerization domain structure." RNA **17**(3): 412-418.

Mosley, A. L., S. G. Pattenden, M. Carey, S. Venkatesh, J. M. Gilmore, L. Florens, J. L. Workman and M. P. Washburn (2009). "Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation." Mol Cell **34**(2): 168-178.

Moucadel, V., F. Lopez, T. Ara, P. Benech and D. Gautheret (2007). "Beyond the 3' end: experimental validation of extended transcript isoforms." Nucleic Acids Res **35**(6): 1947-1957.

Nedea, E., X. He, M. Kim, J. Pootoolal, G. Zhong, V. Canadien, T. Hughes, S. Buratowski, C. L. Moore and J. Greenblatt (2003). "Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3'-ends." J Biol Chem **278**(35): 33000-33010.

Nemeroff, M. E., S. M. Barabino, Y. Li, W. Keller and R. M. Krug (1998). "Influenza virus NS1 protein interacts with the cellular 30 kDa subunit of CPSF and inhibits 3' end formation of cellular pre-mRNAs." Mol Cell **1**(7): 991-1000.

Nevins, J. R. and J. E. Darnell, Jr. (1978). "Steps in the processing of Ad2 mRNA: poly(A)⁺ nuclear sequences are conserved and poly(A) addition precedes splicing." Cell **15**(4): 1477-1493.

Ni, Z., A. Saunders, N. J. Fuda, J. Yao, J. R. Suarez, W. W. Webb and J. T. Lis (2008). "P-TEFb is critical for the maturation of RNA polymerase II into productive elongation in vivo." Mol Cell Biol **28**(3): 1161-1170.

Ni, Z., B. E. Schwartz, J. Werner, J. R. Suarez and J. T. Lis (2004). "Coordination of transcription, RNA processing, and surveillance by P-TEFb kinase on heat shock genes." Mol Cell **13**(1): 55-65.

Noble, C. G., B. Beuth and I. A. Taylor (2007). "Structure of a nucleotide-bound Clp1-Pcf11 polyadenylation factor." Nucleic Acids Res. **35**(1): 87-99.

Noble, C. G., D. Hollingworth, S. R. Martin, V. Ennis-Adeniran, S. J. Smerdon, G. Kelly, I. A. Taylor and A. Ramos (2005). "Key features of the interaction between Pcf11 CID and RNA polymerase II CTD." Nat. Struct. Mol. Biol. **12**(2): 144-151.

Noble, C. G., P. A. Walker, L. J. Calder and I. A. Taylor (2004). "Rna14-Rna15 assembly mediates the RNA-binding capability of *Saccharomyces cerevisiae* cleavage factor IA." Nucleic Acids Res. **32**(11): 3364-3375.

Nunes, N. M., W. Li, B. Tian and A. Furger (2010). "A functional human Poly(A) site requires only a potent DSE and an A-rich upstream sequence." Embo J **29**(9): 1523-1536.

Ohnacker, M., S. M. Barabino, P. J. Preker and W. Keller (2000). "The WD-repeat protein pfs2p bridges two essential factors within the yeast pre-mRNA 3'-end-processing complex." Embo J. **19**(1): 37-47.

Orkin, S. H., T. C. Cheng, S. E. Antonarakis and H. H. Kazazian, Jr. (1985). "Thalassemia due to a mutation in the cleavage-polyadenylation signal of the human beta-globin gene." Embo J **4**(2): 453-456.

Orozco, I. J., S. J. Kim and H. G. Martinson (2002). "The poly(A) signal, without the assistance of any downstream element, directs RNA polymerase II to pause in vivo and then to release stochastically from the template." J Biol Chem **277**(45): 42899-42911.

Osheim, Y. N., N. J. Proudfoot and A. L. Beyer (1999). "EM visualization of transcription by RNA polymerase II: downstream termination requires a poly(A) signal but not transcript cleavage." Mol Cell **3**(3): 379-387.

Pabo, C. O., E. Peisach and R. A. Grant (2001). "Design and selection of novel Cys2His2 zinc finger proteins." Annu Rev Biochem **70**: 313-340.

Pancevac, C., D. C. Goldstone, A. Ramos and I. A. Taylor (2010). "Structure of the Rna15 RRM, ÅRNA complex reveals the molecular basis of GU specificity in transcriptional 3', Å≤-end processing factors." Nucleic Acids Research **38**(9): 3119-3132.

Paulson, A. R. and L. Tong (2012). "Crystal structure of the Rna14-Rna15 complex." RNA **18**(6): 1154-1162.

Peart, N., A. Sataluri, D. Baillat and E. J. Wagner (2013). "Non-mRNA 3' end formation: how the other half lives." Wiley Interdiscip Rev RNA.

Perales, R. and D. Bentley (2009). ""Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions." Mol Cell **36**(2): 178-191.

Perez-Canadillas, J. M. (2006). "Grabbing the message: structural basis of mRNA 3'UTR recognition by Hrp1." Embo J. **25**(13): 3167-3178.

Perez Canadillas, J. M. and G. Varani (2003). "Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein." Embo J. **22**(11): 2821-2830.

Peterlin, B. M. and D. H. Price (2006). "Controlling the elongation phase of transcription with P-TEFb." Mol Cell **23**(3): 297-305.

Petoukhov, M. V., D. Franke, A. V. Shkumatov, G. Tria, A. G. Kikhney, M. Gajda, C. Gorba, H. D. T. Mertens, P. V. Konarev and D. I. Svergun (2012). "New developments in the ATSAS program package for small-angle scattering data analysis." Journal of Applied Crystallography **45**(2): 342-350.

Petoukhov, M. V. and D. I. Svergun (2005). "Global rigid body modeling of macromolecular complexes against small-angle scattering data." Biophys J **89**(2): 1237-1250.

Phatnani, H. P. and A. L. Greenleaf (2006). "Phosphorylation and functions of the RNA polymerase II CTD." Genes Dev **20**(21): 2922-2936.

Preker, P. J., J. Lingner, L. Minvielle-Sebastia and W. Keller (1995). "The FIP1 gene encodes a component of a yeast pre-mRNA polyadenylation factor that directly interacts with poly(A) polymerase." Cell **81**(3): 379-389.

Proudfoot, N. (2004). "New perspectives on connecting messenger RNA 3' end formation to transcription." Curr. Opin. Cell. Biol. **16**(3): 272-278.

Proudfoot, N. J. and G. G. Brownlee (1976). "3' non-coding region sequences in eukaryotic messenger RNA." Nature **263**(5574): 211-214.

Proudfoot, N. J., A. Furger and M. J. Dye (2002). "Integrating mRNA processing with transcription." Cell **108**(4): 501-512.

Puig, O., F. Caspary, G. Rigaut, B. Rutz, E. Bouveret, E. Bragado-Nilsson, M. Wilm and B. Seraphin (2001). "The tandem affinity purification (TAP) method: a general procedure of protein complex purification." Methods **24**(3): 218-229.

Qu, X., J. M. Perez-Canadillas, S. Agrawal, J. De Baecke, H. Cheng, G. Varani and C. Moore (2007). "The C-terminal domains of vertebrate CstF-64 and its yeast orthologue Rna15 form a new structure critical for mRNA 3'-end processing." J. Biol. Chem. **282**(3): 2101-2115.

Raabe, T., K. G. Murthy and J. L. Manley (1994). "Poly(A) polymerase contains multiple functional domains." Mol Cell Biol **14**(5): 2946-2957.

Richard, P. and J. L. Manley (2009). "Transcription termination by nuclear RNA polymerases." Genes Dev **23**(11): 1247-1269.

Rigo, F., A. Kazerouninia, A. Nag and H. G. Martinson (2005). "The RNA tether from the poly(A) signal to the polymerase mediates coupling of transcription to cleavage and polyadenylation." Mol Cell **20**(5): 733-745.

Rigo, F. and H. G. Martinson (2008). "Functional coupling of last-intron splicing and 3'-end processing to transcription in vitro: the poly(A) signal couples to splicing before committing to cleavage." Mol Cell Biol **28**(2): 849-862.

Rigo, F. and H. G. Martinson (2009). "Polyadenylation releases mRNA from RNA polymerase II in a process that is licensed by splicing." RNA **15**(5): 823-836.

Rondon, A. G., H. E. Mischo, J. Kawauchi and N. J. Proudfoot (2009). "Fail-safe transcriptional termination for protein-coding genes in *S. cerevisiae*." Mol Cell **36**(1): 88-98.

Rosonina, E., S. Kaneko and J. L. Manley (2006). "Terminating the transcript: breaking up is hard to do." Genes Dev **20**(9): 1050-1056.

Russnak, R., K. W. Nehrke and T. Platt (1995). "REF2 encodes an RNA-binding protein directly involved in yeast mRNA 3'-end formation." Mol Cell Biol **15**(3): 1689-1697.

Russo, P., W. Z. Li, Z. Guo and F. Sherman (1993). "Signals that produce 3' termini in CYC1 mRNA of the yeast *Saccharomyces cerevisiae*." Mol Cell Biol **13**(12): 7836-7849.

Russo, P., W. Z. Li, D. M. Hampsey, K. S. Zaret and F. Sherman (1991). "Distinct cis-acting signals enhance 3' endpoint formation of CYC1 mRNA in the yeast *Saccharomyces cerevisiae*." EMBO J **10**(3): 563-571.

Ryan, K., O. Calvo and J. L. Manley (2004). "Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease." Rna **10**(4): 565-573.

Sachs, A. and E. Wahle (1993). "Poly(A) tail metabolism and function in eucaryotes." J Biol Chem **268**(31): 22955-22958.

Sadowski, M., B. Dichtl, W. Hubner and W. Keller (2003). "Independent functions of yeast Pcf11p in pre-mRNA 3' end processing and in transcription termination." Embo J **22**(9): 2167-2177.

Schaffer, A. E., V. R. Eggens, A. O. Caglayan, M. S. Reuter, E. Scott, N. G. Coufal, J. L. Silhavy, Y. Xue, H. Kayserili, K. Yasuno, R. O. Rosti, M. Abdellateef, C. Caglar, P. R. Kasher, J. L. Cazemier, M. A. Weterman, V. Cantagrel, N. Cai, C. Zweier, U. Altunoglu, N. B. Satkin, F. Aktar, B. Tuysuz, C. Yalcinkaya, H. Caksen, K. Bilguvar, X. D. Fu, C. R. Trotta, S. Gabriel, A. Reis, M. Gunel, F. Baas and J. G. Gleeson (2014). "CLP1 Founder Mutation Links tRNA Splicing and Maturation to Cerebellar Development and Neurodegeneration." Cell **157**(3): 651-663.

Schrieck, A., A. D. Easter, S. Etzold, K. Wiederhold, M. Lidschreiber, P. Cramer and L. A. Passmore (2014). "RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7." Nat Struct Mol Biol.

Shatkin, A. J. and J. L. Manley (2000). "The ends of the affair: capping and polyadenylation." Nat Struct Biol **7**(10): 838-842.

Sherman, F. (2005). "The importance of mutation, then and now: studies with yeast cytochrome c." Mutat Res **589**(1): 1-16.

Shi, Y., D. C. Di Giammartino, D. Taylor, A. Sarkeshik, W. J. Rice, J. R. Yates, 3rd, J. Frank and J. L. Manley (2009). "Molecular architecture of the human pre-mRNA 3' processing complex." Mol Cell **33**(3): 365-376.

Shuman, S. (2001). "Structure, mechanism, and evolution of the mRNA capping apparatus." Prog Nucleic Acid Res Mol Biol **66**: 1-40.

Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson and D. G. Higgins (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.

Simonsen, C. C. and A. D. Levinson (1983). "Analysis of processing and polyadenylation signals of the hepatitis B virus surface antigen gene by using simian virus 40-hepatitis B virus chimeric plasmids." Mol Cell Biol **3**(12): 2250-2258.

Skaar, D. A. and A. L. Greenleaf (2002). "The RNA polymerase II CTD kinase CTDK-I affects pre-mRNA 3' cleavage/polyadenylation through the processing component Pti1p." Mol Cell **10**(6): 1429-1439.

Soding, J., A. Biegert and A. N. Lupas (2005). "The HHpred interactive server for protein homology detection and structure prediction." Nucleic Acids Res **33**(Web Server issue): W244-248.

- Sun, Z. W. and M. Hampsey (1996). "Synthetic enhancement of a TFIIB defect by a mutation in SSU72, an essential yeast gene encoding a novel protein that affects transcription start site selection in vivo." Mol Cell Biol **16**(4): 1557-1566.
- Svergun, D. (1992). "Determination of the regularization parameter in indirect-transform methods using perceptual criteria." Journal of Applied Crystallography **25**(4): 495-503.
- Svergun, D. I. (1999). "Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing." Biophys J **76**(6): 2879-2886.
- Takagaki, Y. and J. L. Manley (1994). "A polyadenylation factor subunit is the human homologue of the Drosophila suppressor of forked protein." Nature **372**(6505): 471-474.
- Takagaki, Y. and J. L. Manley (1997). "RNA recognition by the human polyadenylation factor CstF." Mol. Cell. Biol. **17**(7): 3907-3914.
- Takagaki, Y., L. C. Ryner and J. L. Manley (1989). "Four factors are required for 3'-end cleavage of pre-mRNAs." Genes Dev. **3**(11): 1711-1724.
- Tan, D., W. F. Marzluff, Z. Dominski and L. Tong (2013). "Structure of histone mRNA stem-loop, human stem-loop binding protein, and 3'hExo ternary complex." Science **339**(6117): 318-321.
- Tian, B., J. Hu, H. Zhang and C. S. Lutz (2005). "A large-scale analysis of mRNA polyadenylation of human and mouse genes." Nucleic Acids Res **33**(1): 201-212.
- Tian, B., Z. Pan and J. Y. Lee (2007). "Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing." Genome Res **17**(2): 156-165.
- Tran, D. P., S. J. Kim, N. J. Park, T. M. Jew and H. G. Martinson (2001). "Mechanism of poly(A) signal transduction to RNA polymerase II in vitro." Mol Cell Biol **21**(21): 7495-7508.
- Tsai, T. F., M. J. Wu and T. S. Su (1998). "Usage of cryptic splice sites in citrullinemia fibroblasts suggests role of polyadenylation in splice-site selection during terminal exon definition." DNA Cell Biol **17**(8): 717-725.
- Vagner, S., C. Vagner and I. W. Mattaj (2000). "The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing." Genes Dev **14**(4): 403-413.
- Valentini, S. R., V. H. Weiss and P. A. Silver (1999). "Arginine methylation and binding of Hrp1p to the efficiency element for mRNA 3'-end formation." Rna **5**(2): 272-280.
- van Helden, J., M. del Olmo and J. E. Perez-Ortin (2000). "Statistical analysis of yeast genomic downstream sequences reveals putative polyadenylation signals." Nucleic Acids Res **28**(4): 1000-1010.
- Viphakone, N., F. Voisinet-Hakil and L. Minvielle-Sebastia (2008). "Molecular dissection of mRNA poly(A) tail length control in yeast." Nucleic Acids Res **36**(7): 2418-2433.
- Vo, L. T., M. Minet, J. M. Schmitter, F. Lacroute and F. Wyers (2001). "Mpe1, a zinc knuckle protein, is an essential component of yeast cleavage and polyadenylation factor required for the cleavage and polyadenylation of mRNA." Mol Cell Biol **21**(24): 8346-8356.
- Volkov, V. V. and D. I. Svergun (2003). "Uniqueness of ab initio shape determination in small-angle scattering." Journal of Applied Crystallography **36**(3 Part 1): 860-864.

- Wang, E. T., R. Sandberg, S. Luo, I. Khrebtkova, L. Zhang, C. Mayr, S. F. Kingsmore, G. P. Schroth and C. B. Burge (2008). "Alternative isoform regulation in human tissue transcriptomes." Nature **456**(7221): 470-476.
- Wang, S. W., K. Asakawa, T. Z. Win, T. Toda and C. J. Norbury (2005). "Inactivation of the pre-mRNA cleavage and polyadenylation factor Pfs2 in fission yeast causes lethal cell cycle defects." Mol Cell Biol **25**(6): 2288-2296.
- Waterhouse, A. M., J. B. Procter, D. M. A. Martin, M. Clamp and G. J. Barton (2009). "Jalview Version 2—a multiple sequence alignment editor and analysis workbench." Bioinformatics **25**(9): 1189-1191.
- West, S., N. Gromak and N. J. Proudfoot (2004). "Human 5' → 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites." Nature **432**(7016): 522-525.
- West, S. and N. J. Proudfoot (2008). "Human Pcf11 enhances degradation of RNA polymerase II-associated nascent RNA and transcriptional termination." Nucleic Acids Res **36**(3): 905-914.
- Whitelaw, E. and N. Proudfoot (1986). "Alpha-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3' end processing in the human alpha 2 globin gene." Embo J **5**(11): 2915-2922.
- Wilusz, J. E. and D. L. Spector (2010). "An unexpected ending: noncanonical 3' end processing mechanisms." Rna **16**(2): 259-266.
- Wu, C. and J. C. Alwine (2004). "Secondary Structure as a Functional Feature in the Downstream Region of Mammalian Polyadenylation Signals." Molecular and Cellular Biology **24**(7): 2789-2796.
- Xiang, K., J. L. Manley and L. Tong (2012). "An unexpected binding mode for a Pol II CTD peptide phosphorylated at Ser7 in the active site of the CTD phosphatase Ssu72." Genes Dev **26**(20): 2265-2270.
- Xiang, K., T. Nagaike, S. Xiang, T. Kilic, M. M. Beh, J. L. Manley and L. Tong (2010). "Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex." Nature.
- Xiang, K., L. Tong and J. L. Manley (2014). "Delineating the structural blueprint of the pre-mRNA 3' end processing machinery." Mol Cell Biol.
- Yang, Q., G. M. Gilmartin and S. Doubie (2010). "Structural basis of UGUA recognition by the Nudix protein CFI(m)25 and implications for a regulatory role in mRNA 3' processing." Proc Natl Acad Sci U S A **107**(22): 10062-10067.
- Yang, X. C., I. Sabath, J. Debski, M. Kaus-Drobek, M. Dadlez, W. F. Marzluff and Z. Dominski (2012). "A complex containing the CPSF73 endonuclease and other polyadenylation factors associates with U7 snRNP and is recruited to histone pre-mRNA for 3' end processing." Mol Cell Biol.
- Yonaha, M. and N. J. Proudfoot (1999). "Specific transcriptional pausing activates polyadenylation in a coupled in vitro system." Mol Cell **3**(5): 593-600.
- Zaret, K. S. and F. Sherman (1982). "DNA sequence required for efficient transcription termination in yeast." Cell **28**(3): 563-573.

Zarudnaya, M. I., I. M. Kolomiets, A. L. Potyahaylo and D. M. Hovorun (2003). "Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures." Nucleic Acids Res **31**(5): 1375-1386.

Zhang, H. and J. K. Zhu (2014). "Emerging roles of RNA processing factors in regulating long non-coding RNAs." RNA Biol **11**(7).

Zhang, Y., Y. Kim, N. Genoud, J. Gao, J. W. Kelly, S. L. Pfaff, G. N. Gill, J. E. Dixon and J. P. Noel (2006). "Determinants for dephosphorylation of the RNA polymerase II C-terminal domain by Scp1." Mol Cell **24**(5): 759-770.

Zhang, Z., J. Fu and D. S. Gilmour (2005). "CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11." Genes & Development **19**(13): 1572-1580.

Zhang, Z. and D. S. Gilmour (2006). "Pcf11 is a termination factor in Drosophila that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript." Mol Cell **21**(1): 65-74.

Zhao, J., L. Hyman and C. Moore (1999). "Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis." Microbiol Mol Biol Rev **63**(2): 405-445.

Zhao, J., M. M. Kessler and C. L. Moore (1997). "Cleavage factor II of *Saccharomyces cerevisiae* contains homologues to subunits of the mammalian Cleavage/ polyadenylation specificity factor and exhibits sequence-specific, ATP-dependent interaction with precursor RNA." J Biol Chem **272**(16): 10831-10838.

Zhelkovsky, A., Y. Tacahashi, T. Nasser, X. He, U. Sterzer, T. H. Jensen, H. Domdey and C. Moore (2006). "The role of the Brr5/Ysh1 C-terminal domain and its homolog Syc1 in mRNA 3'-end processing in *Saccharomyces cerevisiae*." RNA **12**(3): 435-445.